

Moving Forward Against Misinformation or Stepping Back? WhatsApp's Forwarded Tag as an Electronically Relayed Information Cue

EDSON C. TANDOC JR.¹

SONNY ROSENTHAL

JEROME YEO

ZOE ONG

TINGTING YANG

SHELLY MALIK

MENGXUE OU

YICHEN ZHOU

JINGWEI ZHENG

HAMKA AFIQ BIN MOHAMED

JOANNE TAN

ZHI XIN LAU

JIA YAO LIM

Nanyang Technological University, Singapore

WhatsApp, currently the leading messaging application in the world with an estimated 2 billion users, introduced the forwarded tag in July 2018. When WhatsApp users send

Edson C. Tandoc Jr.: edson@ntu.edu.sg

Sonny Rosenthal: sonnyrosenthal@ntu.edu.sg

Jerome Yeo: YEOT0020@e.ntu.edu.sg

Zoe Ong: ONGZ0063@e.ntu.edu.sg

Tingting Yang: TINGTING001@e.ntu.edu.sg

Shelly Malik: SHEL0013@e.ntu.edu.sg

Mengxue Ou: MENGXUE001@e.ntu.edu.sg

Yichen Zhou: YICHEN001@e.ntu.edu.sg

Jingwei Zheng: JINGWEI003@e.ntu.edu.sg

Hamka Afiq Bin Mohamed: HAMK0001@e.ntu.edu.sg

Joanne Tan: JTAN332@e.ntu.edu.sg

Zhi Xin Lau: ZLAU006@e.ntu.edu.sg

Jia Yao Lim: limj0184@e.ntu.edu.sg

Date submitted: 2021-06-17

¹ This study was supported by the WhatsApp Research Award. The authors also wish to thank Dai Rui Kai and Geng Bowen for their assistance during the data collection.

Copyright © 2022 (Edson C. Tandoc Jr., Sonny Rosenthal, Jerome Yeo, Zoe Ong, Tingting Yang, Shelly Malik, Mengxue Ou, Yichen Zhou, Jingwei Zheng, Hamka Afiq Bin Mohamed, Joanne Tan, Zhi Xin Lau, and Jia Yao Lim). Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at <http://ijoc.org>.

messages that they received from someone else, a tag appears with the message to indicate it has been forwarded. By alerting receivers that the message was not written by the immediate sender and was merely passed along, the forwarded tag may trigger skepticism of, and efforts to, verify the message contents. But does it? This study conceptualizes the forwarded tag as an electronically relayed information cue (ERIC) and seeks to answer that question using a mixed-methods study conducted in Singapore. This study combines data from an online experiment ($n = 266$) and individual and group interviews ($n = 65$) in a sequential explanatory design. The online experiment found participants rated a WhatsApp message as less credible when it was accompanied by a forwarded tag, whereas the interviews found users associate the forwarded tag with originality and sincerity.

Keywords: cue, heuristic, messaging apps, misinformation, qualitative, WhatsApp

WhatsApp is currently the most used messaging application with an estimated 2 billion users (Porter, 2020). In July 2018, WhatsApp introduced the forwarded tag to alert users when a message they received was not created by the person who had sent it to them. The tag includes the word “forwarded” and one or two arrows to indicate if the message has been forwarded up to five times or more than five times (Dixit, 2019). The intended function of this tag is straightforward: to help users determine quickly whether the message was created by the sender or created by someone other than the sender (Funke, 2018). WhatsApp’s (2021) website claims this “helps slow down the spread of rumors, viral messages, and fake news” (para. 2).

WhatsApp beta-tested the forwarded tags in India, where many individuals rely on the platform not just for communication but also for information. In 2018, several Indian men were killed by village mobs responding to fake news that child kidnappers were roaming their villages (Frayer, 2018). Much of that misinformation was forwarded on WhatsApp, facilitating its rapid spread. Partly in response to those incidents, WhatsApp developed features to slow the spread of rumors, fake news, and other kinds of false information. One of those features is the forwarded tag.

Other social media platforms, such as Facebook and Twitter, have features indicating the number of times a message has been shared or retweeted. However, those kinds of information cues serve a rather different function than WhatsApp’s forwarded tag, at least in their intended purpose and perhaps in how audiences regard them. The number of times a message has been shared on Facebook or Twitter, indicating its popularity, may trigger a bandwagon heuristic (Sundar, 2008). As a result, audiences may tend to like and share those messages more and be more likely to agree with the ideas they convey. This is clearly in contrast to the purpose of the forwarded tag. Whereas popularity cues have received attention in the literature, the idea of a forwarded tag has not been conceptually defined. It is helpful to develop that definition before thinking about its role in message processing and message effects. Thus, we begin with a proposed conceptualization of forwarded tags as a kind of electronically relayed information cue (ERIC). Following that conceptualization, we use the MAIN (modality, agency, interactivity, navigability)

model of online credibility assessment (Sundar, 2008) to theorize about the effects of ERICs on perceived message credibility.

Literature Review

Since its founding in 2009 and acquisition by Facebook in 2014 (Shinal, 2019), WhatsApp has amassed more than 2 billion monthly users worldwide (Porter, 2020). In Singapore, a small but technologically and financially advanced Asian country, about 84% of the population uses WhatsApp (Tandoc, 2019). WhatsApp allows users to send messages and make phone and video calls over the Internet. Compared with semipublic social media platforms like Facebook and Twitter, WhatsApp provides relatively higher levels of privacy. WhatsApp is generally used to connect with one or a few contacts and promises end-to-end message encryption and thus can be categorized as a “private communication” channel (Karapanos, Teixeira, & Gouveia, 2016, p. 889). On Facebook and Twitter, pertinent information about the user is displayed on their profile, whereas on WhatsApp, a user would have to interact with another to acquire the same information. WhatsApp’s near-synchronous and enclosed nature allows users to emulate in-person chats, which tend to be more intimate.

A WhatsApp account is connected to a user’s mobile phone number, which means another user cannot simply look up a person’s name online to connect on WhatsApp. Although some users occasionally receive spam messages from random senders, most of the time they can control which other WhatsApp users can send them messages. This is probably why Karapanos and colleagues (2016) found that WhatsApp is chiefly used to connect with close or strong ties, while Facebook involves more weak-tie connections. WhatsApp also allows users to create or join chat groups, where users can send messages that all group members can read and respond to. For many users, WhatsApp has become a platform for coordinating with friends and family as well as for sharing information, such as news (Pang & Woo, 2020).

But WhatsApp’s important role in information exchange has also facilitated the spread of misinformation. Because of WhatsApp’s closed nature, unlike semipublic social media platforms such as Facebook and Twitter, fake news and misinformation spread through WhatsApp are more difficult to track. WhatsApp (2021) itself claims it is unable to track the content of messages passed through its platform; other social media platforms, such as Facebook, were also initially unwilling to directly respond to the problem of fake news circulating on their spaces (Lien, Lee, & Tandoc, 2021). But the spate of mob killings in India in 2018, triggered by viral fake news messages on WhatsApp, has triggered massive concern about the role of social media and messaging apps in the spread of online falsehoods. This has led WhatsApp to institute a series of interventions to slow down the spread of fake news on its platform. One of these technological interventions is the introduction of the forwarded tag.

Cues and Heuristics

This study is guided by the MAIN model of online credibility, which identifies four online affordances that affect credibility judgments: modality, agency, interactivity, and navigability (Sundar, 2008). The model argues that within each of these broad categories of affordances, there are cues or markers that users encounter. Those cues may trigger heuristics affecting subsequent message processing. Heuristics refer to

"judgmental rules" that individuals use to make decisions. For example, a long message carries length as a cue that can trigger the "length implies strength" heuristic, which may lead an individual to consider the message as credible (Sundar, 2008, p. 74).

Features of the communication channel affect how individuals respond to messages. For example, Lin, Spence, and Lachlan (2016) considered the presence of retweets in one's Twitter profile as a cue that can trigger perceptions of expertise; Westerman, Spence, and Van Der Heide (2012) examined the number of users that a Twitter user is following as a cue that can affect perceived expertise and trustworthiness of the user. These online cues that affect user perceptions have been broadly typologized into two types: self-generated cues, or those generated and controlled by communicators themselves, such as one's profile photo or self-description, and system-generated cues, or those generated by the system and are beyond the direct control of the communicators, such as the number of likes (Shan, 2016; Tong, Van Der Heide, Langwell, & Walther, 2008). Thus, user judgments can be influenced not just by the content of the communication but also by cues generated by the channel of communication. Sundar (2008) suggested such cues and markers gain salience in the digital era because new communication technologies uniquely afford them. Indeed, before the mainstreaming of the Internet, there was not a straightforward way for most individuals to track and display the popularity of the ideas they expressed publicly. Similarly, the Internet and social media have provided new types of system-generated cues, such as the forwarded tag.

The ability to forward messages is not unique to WhatsApp. "Chain letters" that were sent through the post became a thing before the Internet—a piece of snail mail is delivered and the recipient must send a copy of the letter to 10 other people via snail mail or, as the chain message would claim, face years of bad luck, if not death (Rosen, 2016). When SMS technology came, "chain texts" replaced letters but still operated the same way: Recipients must forward the message to several people, or someone in the family would die. Online, one of the earliest iterations of message forwarding came with electronic mail, or e-mail. Forwarding an e-mail message involves creating an exact duplicate of a message that can be modified by the user and sent to other users (Chiu, Hsieh, Kao, & Lee, 2007). A forwarded e-mail can be identified by having "Fwd." in its subject line (Phelps, Lewis, Mobilio, Perry, & Raman, 2004).

Passing messages continued in the social media era. Facebook allows the sharing of posts through its "share" function. Shared posts will indicate the original post (including photos, videos, and URLs) and the original page or user that posted it. Although Instagram does not allow users to directly share posts from other users, unlike Facebook, it has a "send" function that works similarly to WhatsApp's forward function. This function enables users to send another user's post to another user via an Instagram direct message. The receiver of the message can click on a thumbnail to view the original post. Twitter also allows the reposting or forwarding of a post that was originally made by the user or another through its "retweet" function. Retweets are identified by the retweet icon and the username of the retweeter. A user can also retweet someone else's post with her own comment attached to the original tweet, which is preserved in its original form. WeChat, the most popular messaging app in China, also enables users to share information with their contacts through the "send to chat" button (Gan & Wang, 2015). When selecting a single message, the sent message will copy only the original message. However, when more than one message is shared, the sent messages will be sent in a list titled "Chat History." Unlike in WhatsApp, the chat history displays the name of the original source of the sent messages.

These icons across different platforms signal that the messages they accompany have been forwarded from someone else; the appearance of such tags is beyond the direct control of the immediate sender and may be considered nonvolitional (Shan, 2016). Thus, labels signifying forwarding can be considered as system-generated cues (Tong et al., 2008). And yet, WhatsApp's forwarded tag differs from other well-researched examples of system-generated cues, such as popularity cues (e.g., number of likes). Unlike other system-generated cues on other social media platforms, WhatsApp's forwarding function allows the sharing of someone else's message without displaying the details of the original author or sender, making it extremely difficult to track the original creator of the forwarded message if that detail is excluded from the original message. Thus, in WhatsApp's particular context, the role of the forwarded tag becomes more salient in the absence of other original source-related cues. Although the MAIN model assumes that system-generated cues can affect credibility perceptions, such as the number of likes of a post, the impact of forwarded tags on credibility perceptions has not been explored. Thus, in this study we are proposing another type of system-generated cue that can encompass WhatsApp's forwarded tag, which we refer to as ERIC, defined as a system-generated cue that indicates a message was not originally drafted by the immediate sender and is just being passed on from someone else without any alteration. Although studies have examined the impact of other system-generated cues on credibility judgments online (e.g., Lin et al., 2016; Shan, 2016; Tong et al., 2008; Westerman et al., 2012), studies have not examined the impact of WhatsApp's forwarded tag on credibility judgments, despite WhatsApp's (2021) hopes that it would deter the spread of fake news. Therefore, we propose this question:

RQ1: How does WhatsApp's forwarded tag affect the perceived credibility of a WhatsApp message?

Heuristics From Forwarding

Different factors affect the extent to which a message gets forwarded. For example, Suh, Hong, Pirolli, and Chi (2010) identified three major categories that influence sharing behavior: characteristics of the message source; of the user making sense of the message; and naturally, of the message itself. However, although studies have explored what makes users share, retweet, or forward messages, the impact of such behaviors on message receivers has not been thoroughly investigated. How do recipients perceive the act and the content of a forwarded message? In the context of e-mail, Phelps and colleague (2004) found that individuals commonly opened messages only from somebody they knew. However, they would delete the e-mail if they have received it before or if the subject line of the e-mail showed that it had been forwarded multiple times. On Facebook, it is common to see users sharing posts from brands—sometimes this is a paid endeavor—which has given rise to what users call as “brand ambassadors” (Malhotra, Malhotra, & See, 2013; p. 20). In this context, the sharing of posts is viewed positively. Tandoc, Huang, Duffy, Ling, and Kim (2020) found that sharing and receiving news links on social media had also become part of a tacit social exchange among social media users: Receiving news links from a family member or a friend motivates a user to share a news link in return.

Although the forwarding of messages is not new, WhatsApp has made it faster and easier. However, unlike in forwarded e-mails where a recipient can go down and read previous iterations of the e-mail up to its original sender (unless these are deleted before forwarding), a forwarded message on WhatsApp does

not carry any other detail except the message itself and, through the introduction of the forwarded tag, single or double arrows to indicate that the message had been forwarded.

Although the original intention is to signal receivers that they are receiving a message that had been shared by others, the forwarded tag, just like other system-generated cues online, may have intended and unintended uses and understandings. Instagram, for example, removed the “number of likes” that have always accompanied posts on the platform in response to negative and unintended consequences arising from the pressure to get likes, such as unhealthy diets, risky behavior, and even criminal behaviors or accidents (Fitzgerald, 2019); it later brought it back but allowed account owners to hide the number of likes their posts receive (Criddle, 2021). WhatsApp’s forwarded tag is also similar to other system-generated cues, such as Facebook’s number of shares or Twitter’s number of retweets, and yet these shared cues have been understood as popularity cues, indicating a post’s reach and, in some respects, quality. WhatsApp’s forwarded tag was introduced for a different function. Since system-generated cues can affect how users perceive and respond to a message, it is important to examine how users understand a new type of system-generated cue introduced by WhatsApp. Doing so can help to explain the cue’s impact on message credibility perceptions. Therefore, we also ask:

RQ2: How do WhatsApp users understand WhatsApp’s forwarded tag as a system-generated cue?

We answer these research questions using a combination of quantitative and qualitative approaches in a sequential explanatory design (Creswell & Plano Clark, 2011). This design uses a quantitative approach to establish a general pattern or a relationship, then a qualitative approach to explain the pattern or relationship through meaning-making (Ivankova, Creswell, & Stick, 2006). Thus, we address RQ1 using an online experiment and RQ2 using a combination of in-depth interviews and focus group discussions.

Quantitative Approach

Participants

First, we conducted an online survey-based experiment to examine the impact of the forwarded tag on perceived message credibility. This was part of a larger survey focusing on WhatsApp use in Singapore. We used Dynata, an online survey panel provider, to recruit 400 WhatsApp users to participate in the experiment. Of those invited, 266 provided usable responses. The sample was 56% male and ranged in age from 22 to 84 years ($Mdn. = 41$, $M = 43.53$, $SD = 12.91$). Ethnically, the sample was 54% Chinese, 24% Malay, 17% Indian, and 5% other ethnicities. In terms of highest education level attained, 60% had at least an undergraduate university degree, 22% completed junior college or polytechnic, 17% completed secondary school, and 2% completed primary school.

Procedure

The experiment used a mixed factorial design, where ERIC (forwarded tag present vs. absent) was a within-subjects factor and message leaning (pro- vs. anti-homosexuality) was a between-subjects factor. The messages differed in their presentation of an ERIC, where one message included a forwarded tag and

the other did not. We randomized the message order. Since participants saw both conditions (forwarded tag present vs. absent), we also examined whether seeing one condition before the other affected credibility perceptions of the subsequent message, which may indicate figuring out the study manipulation. A series of independent sample t tests showed this was not the case. For the forwarded tag absent condition, message credibility did not vary whether they saw it before ($M = 3.57$, $SD = 1.09$) or after ($M = 3.42$, $SD = 1.09$) seeing the message that carried the forwarded tag, $t(264) = 1.05$, $p = .294$. Similarly, for the forwarded tag present condition, message credibility did not vary whether they saw it before ($M = 3.33$, $SD = 1.21$) or after ($M = 3.35$, $SD = 1.09$) seeing the message without the forwarded tag, $t(264) = -0.08$, $p = .935$. This gives us confidence that no instrument learning occurred.

Participants were told they would evaluate messages about homosexuality, a polarizing issue in Singapore (Detenber, Ho, Neo, Malik, & Cenite, 2013). Thus, the messages presented positions that were either both pro-homosexuality ($n = 157$) or both anti-homosexuality ($n = 109$), which was assigned randomly. The pro- and anti-homosexuality versions of the messages were matched with respect to the core theme, sentence structure, stylization, and word count. This was to minimize superficial message differences, which would create a confound. After exposure to each screenshot, participants rated the perceived credibility of the message it contained.

Independent Variables

As individuals tend to agree more with messages congruent with their preexisting views (Jones & Sugden, 2001), participants reported their attitude toward homosexuality before viewing the experimental stimuli. We measured their attitude using seven items from the short form of the Attitudes Toward Lesbians and Gay Men (ATLG-S; Herek, 1994) scale, which Detenber and colleagues (2013) used in the Singapore context. Response options ranged from 1 (strongly disagree) to 6 (strongly agree), where a higher score indicates a more negative attitude toward lesbians and gay men. The scale had good reliability (Cronbach's $\alpha = .86$), and the average score was above the hypothetical neutral value of 3.5 ($M = 3.97$, $SD = 1.18$), $t(265) = 6.52$, $p < .001$, suggesting a slight anti-homosexual leaning consistent with the findings of Detenber and colleagues (2013). Using median split, we created two groups based on their scores: those who have positive attitudes ($n = 85$) and those with negative attitudes ($n = 181$). We then included both message condition (pro- vs. anti-homosexuality) and attitude (positive vs. negative toward homosexuality) as between-subjects factors.

Dependent Variable

To measure perceived message credibility, we derived six items from prior studies, including one in the Singapore context (Tandoc, 2019). Participants indicated on a Likert scale their agreement that the message "can be trusted," "is accurate," "is fair," "is believable," "doesn't tell the whole story," and "is biased." The last two items were reverse coded. Response options ranged from 1 (strongly disagree) to 7 (strongly agree), where a higher score indicates greater perceived message credibility. The scale had good reliability (Cronbach's $\alpha = .87$), and the average score was below the neutral response option of 4 ($M = 3.41$, $SD = 0.99$), $t(265) = -9.74$, $p < .001$, suggesting overall low perceived message credibility.

Quantitative Results

RQ1 asked about the impact of an ERIC on perceptions of message credibility. To answer this question, we conducted a mixed factorial analysis of variance (ANOVA). The model included the ERIC manipulation (forwarded tag present vs. absent) as a within-subjects factor and message leaning (pro- vs. anti-homosexuality) and attitude toward homosexuality (positive vs. negative) as between-subjects factors. As expected, there was a significant interaction effect between message leaning and participants' attitude toward homosexuality, $F(1, 262) = 6.17, p = .01, \eta^2_p = .023$. Those who have positive attitudes toward homosexuality rated messages that were pro-homosexuality ($M = 3.55, SE = .14$) as more credible than anti-homosexuality messages ($M = 3.01, SE = .18$), while those who have negative attitudes toward homosexuality rated pro-homosexuality messages ($M = 3.39, SE = .10$) as less credible than anti-homosexuality messages ($M = 3.52, SE = .11$).

Finally, the analysis showed a significant main effect of the ERIC manipulation, $F(1, 261) = 4.35, p = .002, \eta^2_p = .036$. Participants rated the message as lower in credibility when the forwarded tag was present ($M = 3.27, SE = .08$) versus when it was absent ($M = 3.47, SE = .07$). However, there were no interaction effects between ERIC and the message leaning or the participants' attitudes toward homosexuality. Therefore, participants perceive a WhatsApp message as less credible when it is accompanied by a forwarded tag. To explain this effect, we followed up the experiment with a qualitative study.

Qualitative Approach

Procedure

We conducted in-depth interviews and focus group discussions (FGDs) involving WhatsApp users in Singapore ($n = 64$). Interviews were first conducted with 18 participants across different age groups: eight millennials (between 22 and 37 years old), six generation X participants (ages 38–53), and four baby boomers (ages 54–72). The interviewees were evenly split in terms of gender. Initial findings from the interviews were used to draft an FGD guide that was used to run five FGDs involving 46 participants, belonging to the millennial generation, of which 31 were male.

Interviews

The semi-structured interviews were guided by a protocol that asked each participant about their WhatsApp use, their understanding of the WhatsApp forwarded tag, and their responses when they come across the tag. The interviews lasted between 25 minutes to an hour and were conducted face-to-face at locations chosen by the interviewees. The interviews were then transcribed verbatim for data analysis. Following initial analysis, the researchers decided to conduct additional FGDs to explore themes more comprehensively. Sundar (2008) had argued that younger generations, who are digital natives, tend to pay more attention to online cues. Thus, we conducted additional focus group discussions with college students.

Focus Group Discussions

Participants were recruited from a large university in Singapore through a recruitment e-mail sent to the university's mass e-mail system. The participants were invited to an FGD lab, where a moderator facilitated the discussion, guided by our FGD protocol. Developed from the interview guide, the FGD protocol also included questions to ask participants about messages they have received on WhatsApp that carried the forwarded tag. The participants were asked to look at their recent WhatsApp messages and share what messages contained the forwarded tag. An elicitation question was also included, where images of the forwarded tag were shown, and the participants were asked to describe what they thought about what was shown to them. The FGDs lasted for about an hour on average and were also transcribed verbatim for analysis.

Analysis

Interview and FGD data were analyzed together using the constant comparative approach to data analysis that originated in the grounded theory tradition but has since been used widely in qualitative studies (Glaser & Strauss, 1967; Tracy, 2013). This approach requires two stages. In the open-coding stage, the researchers independently read the transcripts for an initial soak in the data and then reread them, this time coding line by line when applicable, to label emerging codes. In the axial-coding stage, the researchers individually revisited the codes and grouped them into larger conceptual bins, sensitized by the study's RQ2 that asked about how users understand the forwarded tag. Coding was also guided by sensitizing concepts from the MAIN model (Sundar, 2008), such as cues and heuristics. The conceptual bins were then discussed and deliberated upon in a group session and became the basis for the themes presented in the next section.

Qualitative Results: What the Tag Means

Most study participants do not remember exactly when they started seeing the forwarded tags. A few took notice of the tags only after they were told about this during the interview. Most participants, however, are familiar with the forward function on WhatsApp and have received messages that included a forwarded tag. These forwarded messages were about various topics, such as real news and fake news; family photos and events; religious messages and updates; restaurant and supermarket deals; and schoolwork, such as instructions from professors and assignments.

Although receiving forwarded messages is not new for the participants—and this is true even before WhatsApp—they all agreed that the forwarded tag represented a new type of cue. A few of them talked about how even before the introduction of the tag, they could guess whether a message was not originally written for them—an example would be WhatsApp messages sent by insurance agents. One FGD participant said she could see how the messages from an insurance agent selling her a policy were based on existing and previous messages. But with the forwarded tag on WhatsApp, she said she no longer had to rely on her hunch in determining whether or not a message sent to her was original.

RQ2 asked about the meanings that the participants ascribe to WhatsApp's forwarded tag. Although this study conceptualizes the forwarded tag as an online cue and proposes it as an ERIC, how users interpret it can also be conceptualized as a set of heuristics that an ERIC can trigger (Sundar, 2008). Heuristics,

however, also work subconsciously, so the results presented here refer only to the range of the participants' conscious interpretations of what the forwarded tag means to them. In response to RQ2, we examined how the participants understand the forwarded tag as well as how they respond to it. In terms of their understanding of the cue, the analysis found four themes: originality, reliability, sincerity, and belonging.

Originality

Some of the meanings ascribed to the forwarded tag were consistent with what WhatsApp had intended for the forwarded tag to signal—that the forwarded message was not originally created by the immediate sender. Some participants noted that the forwarded tag signals to them that the sender just passed the message and did not personally create it. A millennial interviewee said:

The minute I see "forwarded," it means to me that this wasn't, he wasn't the original source of the, of the content. . . . So, it could have been from many—like it wouldn't, it will be, somewhere from, you know, outside. It's like outside information.

But other participants understand the sense of originality communicated by the forwarded tag in a different way. Since the immediate sender just passed the message created by someone else, some participants regard a forwarded message to be in its original, unaltered form. That is, the immediate sender did not change or tamper with the message. For example, an FGD participant said, citing a forwarded message he had received:

My research assistant forward the message from my Professor to the group. "This is what you do." Because, like, it seems legitimate because, like, he didn't edit the message to, like, say you need to do more work. Just that he send whatever that the Prof says to do.

Reliability

Some participants also interpreted the forwarded tag as a signal of the message's degree of reliability. For example, an FGD participant referred to the difference between a single arrow and a double arrow. WhatsApp's forwarded tag can come with a single arrow to indicate the message had been forwarded fewer than five times, while a double arrow means the message had been forwarded five or more times away from the original creator or sender. The participant said: "The two arrows one means that it's been forwarded more times, so it's more reliable."

In contrast, other participants noted that the immediate sender's lack of control with the original message that was forwarded signals that the immediate sender is not a reliable source. This addresses the complex sourcing system in social media and messaging app platforms, where an immediate sender can be different from the original source of the message (Kang, Bae, Zhang, & Sundar, 2011). Another FGD participant said:

It's not from whoever you are expecting from. So, it could be a unreliable source because the person just forward the text, it's not the person go and read, then copy and paste the link and paste because the person just forward so it's, I think, just to let you know this is not, this source is not from the person you are contacting from.

Insincerity

Some participants said the forwarded tag signals some degree of insincerity from the immediate sender. They said that the tag tells the user that the sender merely passed an existing message, instead of communicating an original message and including her personal thoughts. An FGD participant said the forwarded tag "shows that the person didn't take that effort to copy and paste, just forward." This differentiation between forwarding and copy-pasting came up in several responses, where participants considered copy-pasting as more effortful than forwarding. A millennial interviewee also said forwarded messages "feel more disconnected."

A female interviewee from the generation X group compared forwarded WhatsApp messages with chain e-mails. She explained that forwarding messages might be acceptable in certain situations, such as in the workplace where information needs to be disseminated, but it could be problematic when the message is one that should have been personalized. In contrast to other participants, she equated merely forwarding a message to copy-pasting:

If it's a piece of information that is coming top down and obviously this information is coming from somewhere and it needs to be forwarded, I don't mind just forwarding so that they know that I'm receiving this maybe like from my boss, and then forwarding down to their respective colleagues and so on. But if it's something personal to my friends, then I wouldn't want to be seen as you know, I am just copy and paste, I'm just conveniently forwarding information to you [without] even sifting it through.

Sense of Belonging

For some participants, receiving a message that carries a forwarded tag also means being part of a conversation, making them feel part of a social group and affirming their social relationships and involvement. For example, a female participant from the baby boomers group mentioned that she often received greetings and news in the form of forwarded messages. Though these messages were forwarded, the participant said she was "always very happy" about receiving them: "I feel happy because they are willing to share information with me." In this case, forwarded messages seemed to be perceived as an indication that the sender had thought about the recipient, and a sign of the closeness of their relationship.

Another interviewee from the baby boomer group also recalled receiving a forwarded message about the results of a charity singing competition that she had helped organize. In receiving the message, the forwarded tag served as an affirmation of her involvement with the event. The forwarded tag signaled to her that she belonged to the group of people deemed relevant to receive updates about the competition.

Some FGD participants also echoed the same feeling, especially in the context of schoolwork. One participant shared about receiving a forwarded message from a senior that contained details about final exams from previous years. The participant surmised that the message had also been forwarded to the senior by previous students. It made the participant feel grateful to be part of the group that is given access to such a resource.

Qualitative Results: How Users Respond to the Tag

When it comes to how the participants respond to the tag when they see it, the analysis found three different themes. First, some participants reported being more cautious with messages that are accompanied by the cue. Second, some participants used the cue to deprioritize messages. Third, some participants also interpreted the forwarded tag along with other cues.

Becoming More Cautious

Some participants reported becoming more cautious when assessing messages that carried the forwarded tag. This is consistent with WhatsApp's original intention. For example, a female participant from the generation X group said:

It's one of the good thing that WhatsApp introduced, because you know that this one is a forwarded message. So, if you think it is important, maybe it is wise to double check the content first, instead of just simply forwarding.

Similarly, an FGD participant compared his experience receiving forwarded messages when no forwarded cues were available with his experience now that forwarded messages come with the forwarded tag. He said that the forwarded tag had made him more careful about the veracity of information being passed around:

I remember got one time, when they didn't have this function, if a relative or whoever send you maybe about like health thing, you might actually think it's real. But with the forwarded message, you know that actually it's a chain message kind of thing. Then you will be more cautious.

Deprioritizing Messages

Some participants tend to attach less importance to messages that carry the forwarded tag. If the immediate sender did not take time to write an original message or to edit an existing message, then the message is probably not that important. Thus, WhatsApp users who get numerous messages from different senders respond to messages with the forwarded tag by considering them as not a priority. For example, a male interviewee from the millennials group said:

Normally if you just send something important to someone else, it won't be a forwarded message. . . . If I want to tell them something important, it would be something that I know from myself, and I will send it as a direct message instead of forwarding it from someone else.

Such deprioritization also makes it easier for recipients to correct or debunk forwarded messages. Knowing that the message was just forwarded by the sender and not the sender's original message, some recipients feel less concerned about correcting the message. Thus, the sense of detachment that a forwarded cue signals can be an advantage for recipients who want to correct or debunk a forwarded fake news post. For example, an FGD participant said:

It makes it less awkward if you were to just say like "it's false." Because if, let's say, now that we say that it's forwarded, so you are, like, "no wonder." So, if you were to, like, say you were to find out that it's not true, it is easier to go and say: "You don't forward anyhow, this one wrong." Or: "This one is false."

Considering Other Cues

Some participants referred to other cues in responding to the forwarded tag as a cue. Depending on the other cues, their response to the forwarded tag would be different. For example, some participants referred to the sender of the message as an important cue. The forwarded tag becomes an alarm signal for messages that come from a sender known to forward pieces of misinformation. Content and context also matter. For example, forwarded messages are considered normal and routine in certain contexts, such as work and school. An FGD participant said: "If it's like school or work-related, then I find it more reliable because it came from a source and cannot edit, because it's forwarded."

Discussion and Conclusion

Guided by the framework of the MAIN model of online credibility (Sundar, 2008), this mixed-methods explanatory study investigated the impact of WhatsApp's forwarded tag on perceived message credibility as well as how users understand the tag as a cue. We also conceptualized the forwarded tag as an ERIC that can trigger heuristics among users that can then affect credibility judgments online, consistent with the assumption of the MAIN model. Through an online experiment, we found that participants rated messages containing the cue as less credible than those without. Through interviews and FGDs, we found that this might be due to the cue's impact on how original and sincere message recipients perceive a forwarded message is.

Digital system-generated cues are important, for they trigger mental heuristics that then influence credibility judgments online (Sundar, 2008). In this study, we focused on ERICs, which now accompany messages being forwarded on various messaging platforms. On WhatsApp, ERIC takes the form of the forwarded tag. Thus, it is also important to understand how WhatsApp users interpret what ERICs mean. Such interpretations form part of the potential heuristics that the forwarded tag as a cue can trigger, which can affect how users assess the WhatsApp messages they receive. In this period where misinformation and

fake news have become widespread, particularly on messaging apps that are more difficult to track, understanding how certain digital cues trigger what types of mental judgmental rules is crucial. This study found four different ways the participants understood the forwarded tag: originality, reliability, insincerity, and belonging.

First, some participants considered the cue as signaling originality. This seems to be consistent with WhatsApp's original intention, which is to alert message recipients that a message was not originally created by the immediate sender. This might also explain why our online experiment found that the presence of the forwarded tag helps reduce perceived message credibility. However, we also found that some participants interpreted the tag to mean that the forwarded message remained in its original form, unaltered by the immediate sender. This *forwarded-means-unaltered* heuristic may be helpful in certain contexts, such as forwarded messages about work schedules or instructions, as the recipients are assured the content has not been altered from being passed around. But it might be problematic when users equate the original message being intact with accuracy of information, such as when the message in question is a piece of misinformation.

Second, some participants considered the cue as signaling reliability. On one hand, some respondents interpret the forwarded tag to be similar to a Facebook share or retweet—that is, the more often a message had been forwarded, the more popular it is, and hence the more reliable it is (i.e., *a message forwarded often is a popular message*). On the other hand, some participants interpret the cue as signaling the unreliability of the immediate sender, who did not care to comment or add to the message and simply passed it around.

Third, the impact of the forwarded tag on the recipient's perception of the immediate sender also manifests in perceived sincerity. Not only does a forwarded message signal unreliability, for some it also shows lack of sincerity, since the user did not take any effort to compose an original message. Thus, ERICs can also generate heuristics related to social relationships and not just to information quality (e.g., *forwarding messages means not exerting any effort*). Finally, and consistent with heuristics related to social relationships, some participants also interpreted the cue as indicating a sense of belonging. The forwarded tag indicates the message has been sent multiple times to multiple people and seeing it on a message informs users they were part of a select group of recipients, giving them a sense of community.

These different meanings ascribed to the forwarded tag as an ERIC can be understood further in relation to how the participants responded to the cue. Some of them use the cue to guide how they prioritize WhatsApp messages they receive, with tagged messages getting less priority. Similarly, some participants also use the cue as a signal to be cautious with the message, which is consistent with WhatsApp's original intention. Finally, the participants respond to the forwarded tag in relation to other cues. A message accompanied by an ERIC may elicit more caution, depending on the immediate sender and the context of the message. For example, participants consider the forwarded tag as a cue that indicates a message's reliability for being unaltered in the context of workplace communication, such as messages that contain company instructions or announcements. This is particularly interesting. Not only do participants pay attention to multiple cues when assessing a message on WhatsApp, but what these cues mean and how salient they are in credibility judgments also seem to vary across contexts and potentially across different

users. Future studies can build on the results presented here in further unpacking how different users navigate different forms of ERIC.

Furthermore, it seems that ERICs can trigger heuristics that are not only about the message *per se*, but also about the message sender. Thus, although the forwarded tag was originally developed to assist users in assessing the message, users might also be using it, consciously or not, in assessing the sender. It will be illuminating to find out how routinely receiving tagged messages from a sender will affect that recipient's perception of the sender over time; this might then affect how the recipient assesses subsequent messages from the same sender. It is important to note that messaging app users are used to receiving forwarded messages (e.g., copy-pasted messages). What ERICs do is make *more salient* the fact that the message had been forwarded. Future studies should examine whether that basic assumption is achieved—does carrying a forwarded tag make the idea of being a forwarded message more salient, or does this additional cue instead distract the recipient?

It is possible that the introduction of forwarded cues, originally aimed at curbing the spread of falsehoods on WhatsApp, can exert a backfire effect on some users. Not only does one notion of originality—where a recipient takes the message with a forwarded cue as one that is preserved in its original form—can lead some users into believing an unvetted message, but the idea of a forwarded tag indicating some level of insincerity and lack of effort can also lead some senders to copy-pasting messages instead of merely forwarding them. This might facilitate the spread of unreliable information without the forwarded tag as an indicator, if senders go around the forwarded cue by copy-pasting questionable messages that can then mislead some of the recipients. Future studies should keep track of this.

We need to cue in some study limitations. First, given the study's exploratory nature, the results here are descriptive, a goal that was clear for the researchers at the onset, as the main objective was to understand how users make sense of a relatively new digital cue on a widely used communication platform. Future studies can seek to map out which types of users are more likely to understand the cue in a particular way. Second, although we sought to include the perspectives of a range of users in Singapore by recruiting a nationally drawn sample for our online experiment and across generations for our interviews, our FGDs involved only college students. This does not mean the perspectives of other generations were completely excluded—indeed our study 2 results also draw from interview responses across different age ranges—but our qualitative data may have brought in more depth in terms of young adults' experiences. Future studies can build on our qualitative results to conduct similar FGDs involving other generations.

Given WhatsApp's position as the leading messaging app, how it responds to the problem of fake news and other forms of misinformation is important. This current study focused on unpacking how users make sense of WhatsApp's introduction of the forwarded tag, an example of what we call in this study an ERIC. Other messaging apps have also instituted different forms of ERICs. Telegram, for example, also indicates that a message had been forwarded while including the original source. Thus, it will not be surprising to see different types of ERICs being rolled out in other digital platforms in the years to come, cues that indicate frequency but are not supposed to trigger bandwagon heuristics. Whether such a cue helps in curbing the spread of online falsehoods depends on how users make sense of it and investigating WhatsApp's experience will be instructive of the promise and perils of relying on this cue.

References

- Chiu, H. C., Hsieh, Y. C., Kao, Y. H., & Lee, M. (2007). The determinants of e-mail receivers' disseminating behaviors on the Internet. *Journal of Advertising Research*, 47(4), 524–534. doi:10.2501/S0021849907070547
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. Thousand Oaks, CA: SAGE Publications.
- Criddle, C. (2021, May 26). Instagram lets users hide likes to reduce social media pressure. *BBC News*. Retrieved from <https://www.bbc.com/news/technology-57254488>
- Detenber, B. H., Ho, S. S., Neo, R. L., Malik, S., & Cenite, M. (2013). Influence of value predispositions, interpersonal contact, and mediated exposure on public attitudes toward homosexuals in Singapore. *Asian Journal of Social Psychology*, 16(3), 181–196. doi:10.1111/ajsp.12006
- Dixit, P. (2019, August 2). WhatsApp is now letting users know when a message has been forwarded too many times. *BuzzFeed News*. Retrieved from <https://www.buzzfeednews.com/article/pranavdixit/whatsapp-double-arrows-forwarded-messages-misinformation>
- Fitzgerald, M. (2019, July 18). Instagram starts test to hide number of likes posts receive for users in 7 countries. *Time*. Retrieved from <https://time.com/5629705/instagram-removing-likes-test>
- Fraye, L. (2018, July 18). Viral Whatsapp messages are triggering mob killings in India. *NPR*. Retrieved from <https://www.npr.org/2018/07/18/629731693/fake-news-turns-deadly-in-india>
- Funke, D. (2018, July 10). WhatsApp launches a feature that labels forwarded messages. *Poynter*. Retrieved from <https://www.poynter.org/fact-checking/2018/whatsapp-launches-a-feature-that-labels-forwarded-messages>
- Gan, C., & Wang, W. (2015). Uses and gratifications of social media: A comparison of microblog and WeChat. *Journal of Systems and Information Technology*, 17(4), 351–363. doi:10.1108/JSIT-06-2015-0052
- Glaser, B. G., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. London, UK: Weidenfeld and Nicolson.
- Herek, G. M. (1994). Assessing heterosexuals' attitudes toward lesbians and gay men: A review of empirical research with the ATLG scale. In B. Greene & G. M. Herek (Eds.), *Lesbian and gay psychology: Theory, research, and clinical applications* (pp. 206–228). Thousand Oaks, CA: SAGE Publications. doi:10.4135/9781483326757.n11

- Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods*, 18(1), 3–20. doi:10.1177/1525822X05282260
- Jones, M., & Sugden, R. (2001). Positive confirmation bias in the acquisition of information. *Theory and Decision*, 50(1), 59–99. doi:10.1023/a:1005296023424
- Kang, H., Bae, K., Zhang, S., & Sundar, S. S. (2011). Source cues in online news: Is the proximate source more powerful than distal sources? *Journalism & Mass Communication Quarterly*, 88(4), 719–736. doi:10.1177/107769901108800403
- Karapanos, E., Teixeira, P., & Gouveia, R. (2016). Need fulfillment and experiences on social media: A case on Facebook and WhatsApp. *Computers in Human Behavior*, 55(1), 888–897. doi:10.1016/j.chb.2015.10.015
- Lien, C. H., Lee, J., & Tandoc, E. (2021). Facing fakes: Understanding tech platforms' responses to online falsehoods. *Digital Journalism*, 1–20. doi:10.1080/21670811.2021.1982398
- Lin, X., Spence, P. R., & Lachlan, K. A. (2016). Social media and credibility indicators: The effect of influence cues. *Computers in Human Behavior*, 63, 264–271. doi:10.1016/j.chb.2016.05.002
- Malhotra, A., Malhotra, C. K., & See, A. (2013). How to create brand engagement on Facebook. *MIT Sloan Management Review*, 54(2), 18–20. Retrieved from <https://sloanreview.mit.edu/article/how-to-create-brand-engagement-on-facebook>
- Pang, N., & Woo, Y. T. (2020). What about WhatsApp? A systematic review of WhatsApp and its role in civic and political engagement. *First Monday*, 25(12). doi:10.5210/fm.v25i12.10417
- Phelps, J. E., Lewis, R., Mobilio, L., Perry, D., & Raman, N. (2004). Viral marketing or electronic word-of-mouth advertising: Examining consumer responses and motivations to pass along e-mail. *Journal of Advertising Research*, 44(4), 333–348. doi:10.1017/S0021849904040371
- Porter, J. (2020, February 12). WhatsApp now has 2 billion users. *The Verge*. Retrieved from <https://www.theverge.com/2020/2/12/21134652/whatsapp-2-billion-monthly-active-users-encryption-facebook>
- Rosen, J. (2016, October 31). A brief history of the chain letter. *Mental Floss*. Retrieved from <https://www.mentalfloss.com/article/87625/brief-history-chain-letter>
- Shan, Y. (2016). How credible are online product reviews? The effects of self-generated and system-generated cues on source credibility evaluation. *Computers in Human Behavior*, 55, 633–641. doi:10.1016/j.chb.2015.10.013

- Shinal, J. (2018, January 19). Jan Koum got the idea for \$19 billion WhatsApp after missing too many iPhone calls at the gym. *cnbc.com*. Retrieved from <https://www.cnbc.com/2018/01/19/how-jan-koum-got-the-idea-for-whatsapp.html>
- Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010). Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. In *Proceedings of the 2010 IEEE Second International Conference on Social Computing* (pp. 177–184). Washington, DC: IEEE. doi:10.1109/SocialCom.2010.33
- Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 73–100). Cambridge, MA: The MIT Press.
- Tandoc, E. C. Jr. (2019). Singapore. *Digital News Report*. Retrieved from <http://www.digitalnewsreport.org/survey/2019/singapore-2019>
- Tandoc, E., Huang, A., Duffy, A., Ling, R., & Kim, N. (2020). To share is to receive: News as social currency for social media reciprocity. *Journal of Applied Journalism & Media Studies*, 9(1), 3–20. doi:10.1386/ajms_00008_1
- Tong, S. T., Van Der Heide, B., Langwell, L., & Walther, J. B. (2008). Too much of a good thing? The relationship between number of friends and interpersonal impressions on Facebook. *Journal of Computer-Mediated Communication*, 13(3), 531–549. doi:10.1111/j.1083-6101.2008.00409.x
- Tracy, S. (2013). *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact*. Malden, MA: Wiley-Blackwell.
- Westerman, D., Spence, P. R., & Van Der Heide, B. (2012). A social network as information: The effect of system generated reports of connectedness on credibility on Twitter. *Computers in Human Behavior*, 28(1), 199–206. doi:10.1016/j.chb.2011.09.001
- WhatsApp. (2021). *About forwarding limits*. FAQ. Retrieved from <https://faq.whatsapp.com/general/chats/about-forwarding-limits>