# PERSONALIZED INTERACTION-FOCUSED INTERVENTIONS FOR MITIGATING MISINFORMATION

by

Safat Siddiqui

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Software and Information Systems

Charlotte

2022

Approved by:

Dr. Mary Lou Maher

Dr. Heather Lipford

Dr. Wenwen Dou

Dr. Doug Markant

Dr. Jeanette Bennett

# ABSTRACT

SAFAT SIDDIQUI. Personalized Interaction-Focused Interventions for Mitigating Misinformation. (Under the direction of DR. MARY LOU MAHER)

This dissertation addresses a novel approach to assessing users' interaction tendencies on social media as a basis for personalized interventions that can make the truth louder and mitigate the spread of misinformation. This research leverages users' high and low interaction tendencies to amplify truth by increasing users' interactions with verified posts and decreasing their interactions with unverified posts. For designing personalized interaction-focused interventions, this dissertation presents an Active-Passive (AP) framework and three principles of social media interactions to make the truth louder on social media. This dissertation presents a study including tasks and questionnaires to investigate users' differences in the Active-Passive (AP) framework for utilizing platforms' basic interaction functionalities, such as like, comment, or share buttons. The results show that users use the interaction functionalities differently due to their interaction tendencies; users with high interaction tendencies use more interaction functionalities, and users with low interaction tendencies use less.

This dissertation presents an analysis of participants' responses to the design principles and investigates users' additional sharing functionality usage and preference for platform-based incentives. The results show that users with lower interaction tendencies share verified information more when they receive additional interaction support. Furthermore, due to the interaction tendencies, users exhibit opposite preferences for platform-based incentives that can encourage their participation in making the truth louder. Users with high interaction tendencies prefer incentives that highlight their presence on the platform, and users with low interaction tendencies favor incentives that can educate them about the impact of their participation on their friends and community. This dissertation concludes with a discussion on personalized interaction-

focused interventions and provides directions for future research.

# DEDICATION

My father, Nurul Kabir Siddiqui, the most excited person regarding my doctorate degree, used to call me Safat Siddiqui, Ph.D., from an early age in my life. This dissertation is dedicated to my father.

My mother, Kulsuma Begum Shelon, passionately believes that I can contribute to the well-being of society. This dissertation is dedicated to my mother.

# ACKNOWLEDGEMENTS

I am grateful to my advisor, Dr. Mary Lou Maher. Without Dr. Maher's genuine interest and consistent support, this dissertation would not see the light of success.

I am thankful to my committee members Dr. Heather Lipford, Dr. Wenwen Dou, Dr. Doug Markant, and Dr. Jeanette Bennett, for their cordial support and valuable feedback on this dissertation research.

I am thankful to my wife, Kazi Rifat Antara, for providing the necessary support during the challenging time of my Ph.D. journey.

I would like to thank my friends at UNC Charlotte for being available to me for research input and discussions.

Finally, I am grateful to the Almighty God for all kinds of support that were visible or invisible to me.

TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

AP Framework An acronym for Active-Passive Framework.

FBM An acronym for Fogg Behavior Model.

# CHAPTER 1: INTRODUCTION

The spread of misinformation - stories posted on social platforms but unverified or questionable, is responsible for increasing polarization and the consequential loss of trust in science and media [3]. Platforms such as Facebook put related fact-checked article sections to mitigate the effect of questionable content [4]; Twitter warns users about the harmful tweets on their platform [5]. In addition to changes in the social media platforms, providing users with different viewpoints and credibility indicators is a research topic [6, 7, 8, 9, 10]. These platform-based interventions primarily address the factual position of the content and focus on assisting users in making informed decisions regarding their interactions with content. However, these interventions do not address users' various interaction habits on social platforms or leverage their interaction tendencies to combat misinformation. Hence, the existing platform-based interventions do not incorporate a personalized approach to appeal to users with various interaction tendencies to increase participation in combating misinformation.

This dissertation pursues a novel research direction that addresses users' interaction tendencies as a basis for personalized interventions and directs users' interactions to amplify truth for mitigating misinformation. This research begins with developing a theoretical framework from the existing literature to understand social media users' interaction tendencies and usage patterns. The framework prepares the foundation for developing three principles of social media interactions that incorporate the difference in users' interaction tendencies and direct their interactions to promote essential social media behaviors to amplify truth. To examine the implications of the three principles, we design a survey study and investigate users' responses to the design principles. This dissertation presents the significance of the three principles for de-

veloping personalized interventions that combat misinformation by making the truth louder.

## 1.1 Research Motivation

Social media users' interaction tendencies can be described as a continuum from active to passive. Active users tend to produce digital footprints [11, 12], whereas passive users consume content [13, 12]; such patterns have different impacts on users' usage preferences [12]. However, the previous research does not investigate the effect of the interventions on users with various social media usage preferences. For example, Yaqub et al. [6] have studied the effect of credibility indicators on users, Geeng et al. [14] have investigated how users interact with misinformation. These studies have not identified users' interaction tendencies in the active-passive continuum and correlated their findings in response to users' active-passive tendencies. This study distinguishes users based on their active-passive tendencies and investigates their differences in utilizing the basic interaction functionalities such as like, comment, and share buttons. This study investigates how the perception of content impacts the usage of interaction functionalities for users with different interaction tendencies.

Despite the ongoing development of sophisticated algorithms, misleading information is still posted and spread on the platforms. Researchers have investigated effective ways of correcting the misinformation that has already spread, and identified the negative effects of fake information on individuals due to cognitive biases, such as confirmation bias, continued influence, backfire effect [3, 15, 16]. The platform-based interventions create indicators that assist users in making informed decisions on their choices of information consumption and information sharing on the platform. The primary focus of existing design interventions is to communicate to users about the credibility of the content. However, this research shifts the focus to creating intervention designs that consider the difference between active and passive users and are adaptive to individuals' interaction tendencies. Preece et al. [17] have suggested

the importance of various interface supports to increase participation more generally, whereas the dissertation focuses on increasing participation to make the truth louder on social platforms and mitigate the spread of misinformation.

This dissertation develops three principles of social media interaction to make the truth louder that address users' interaction tendencies in addition to the factual position of the content. The three principles: awareness, guidance, and incentive, fulfill specific design purposes [1]. The awareness principle focuses on the design goal that raises users' contextual awareness about the information; the guidance principle focuses on the design goal that provides support facilitating users' interactions. Finally, the incentive principle focuses on the design goal that increases users' encouragement. These principles structure the design space around misinformation and enable designers to consider users' active-passive tendencies for developing personalized platform-based interventions.

This thesis investigates the effect of interaction design on the user across the active-passive continuum when the design adopts the awareness principle and provides factual information about the content. Similarly, this study investigates the effect of a design on the users across the active-passive spectrum that adopts the guidance principle and provides additional interaction support to increase users' participation in combating misinformation. This research is not a UI contribution; instead, the UI is used to study the responses of users to the design principles for personalized interaction-focused intervention. Finally, this research explores how the users across the active-passive continuum show preferences for the platform-based incentives that a design adopting the incentive design principle can utilize. The findings of these investigations can identify the effective design principles that can be applied to develop personalized interaction-focused interventions to amplify the truth on social platforms.

## 1.2 Thesis Statement

The thesis statement is:

*Users across the active-passive continuum have different interaction tendencies, and personalized interaction-focused interventions can leverage the tendencies to increase users' participation in mitigating misinformation by making the truth louder on social platforms.*

## 1.3 Research Questions

The four research questions this dissertation aims to answer are:

- **RQ1:** How do users across the active-passive continuum use the basic interaction functionalities (like, comment, share) differently when they are presented with the content's factual position (awareness design principle)?
- **RQ2:** How do users' perceptions of content influence their usage of basic interaction functionalities across the active-passive continuum?
- **RQ3:** How do users across the active-passive continuum participate in combating misinformation when they are provided with additional interaction functionalities (guidance design principle)?
- **RQ4:** How do users across the active-passive continuum have preferences for the platform-based incentives that can increase their participation in combating misinformation (incentive design principle)?

## 1.4 Methodology

We have designed a survey study to find answers to the four research questions. The survey has 2 experiment conditions: control condition and treatment condition. The control condition studies users' interactions with the verified and unverified posts that adopt the awareness principle. Similarly, the treatment condition examines users'

participation in combating misinformation when the design adopts the guidance principle. In addition, the survey includes a social media usage questionnaire to determine participants' position in the active-passive continuum and an incentive-related questionnaire to investigate participants' preference for platform-based incentives. More details about the study design, participant recruitment, and data analysis can be found in Chapter 5.

## 1.5 Contributions

The contributions of this thesis are as follows:

1. This dissertation investigates a novel research direction that addresses users' interaction tendencies as the basis of personalized interventions for transforming users' social media interaction behaviors to combat misinformation.

2. This dissertation develops a theoretical framework, the Active-Passive (AP) framework, to distinguish users' active-passive interaction tendencies and usage preferences that have emerged from the literature review of users' social media interactions, usage, and behaviors.

3. This dissertation develops three principles of designing social media interactions that leverage users' interaction tendencies to promote necessary interaction behaviors that amplify truth and mitigate misinformation.

4. This dissertation designs a survey study that investigates users' active-passive tendencies, usage of interaction functionalities, and preferences of platform-based incentives to examine the implications of the three principles for developing personalized interaction-focused interventions.

## 1.6 Thesis Overview

Chapter 1 provides the introduction of the thesis.

Chapter 2 presents the relevant literature review on combating misinformation and identifies a gap in the personalized intervention that addresses users' interaction tendencies.

Chapter 3 develops a theoretical framework to understand the difference between users due to their interaction tendencies and social media usage preferences.

Chapter 4 develops three principles of social media interactions that enable designers to explore the information space and design personalized interaction-focused interventions.

Chapter 5 presents the experimental design conducted to investigate users' responses to three design principles and find answers to the four research questions.

Chapter 6 presents the effect of 3 principles on users with different interaction tendencies and shows experiment results and answers to the four research questions.

Chapter 7 discusses the personalized interaction-focused interventions and concludes the dissertation with a conclusion about the thesis.

# CHAPTER 2: RELEVANT RESEARCH ON FIGHTING MISINFORMATION

Fighting misinformation demands interdisciplinary research efforts that include a combined effort of fact-checking services, media literacy approaches, and effective nudging prompts. This chapter presents three directions of research that focus on combating fake news: 1. Fact-checking services, 2. Media literacy, and 3. Nudging prompts and platform-based interventions. The Fact-checking service focuses on detecting fake content and identifying the sources that spread fake news. Media literacy aims to educate and communicate with individuals about the correct information. The nudging prompts and platform-based interventions use visual cues and prompt to draw users' attention to information accuracy and source credibility.

## 2.1 Fact-Checking Services

This section presents an overview of fact-checking services as the research uses the service to detect information veracity. Automated fact-check algorithms can identify fake news, rumor, hoax, and separate legitimate content. There are different ways of detecting fake news that can be categorized into three types [18]: 1. Text-based, 2. Graph-based, and 3. Model-based. Textual-feature based algorithms analyze the textual features of false and real news and identify the differences to build classifiers. Graph-based algorithms detect false information by analyzing the network structures, and model-based algorithms detect those by creating information propagation models.

### 2.1.1 Text-Based Detection

In this section, we present different text-based features that are identified to build classifiers for detecting fake news and discuss the findings of the text based approaches.

Textual analysis of fake news and real news show that these two types are significantly different in style, language choice, and title structure. Horne et al. [19] use Buzzfeed election dataset, Burfoot and Baldwin dataset, and their collected political news dataset, and categorize the content-based features into three broad categories:stylistic features, complexity features, and psychological features.

The producers of fake news capture the attention of the consumers who like to get an overview of the content by reading the headline. Those consumers are less likely to read the whole content thoroughly to understand the arguments of the content. The headlines of the fake news capture the main claims into the titles so that consumers can skip reading the article [19]. For that reason, headlines of fake news are extended, use few stop words, and contain claims related to a specific person and entity. But the content of the fake news tends to be short, repetitive, and less informative. Horne et al. [19] find that fake news is more similar to satire than to real news. Both fake and satire news are written in a less investigative way. Fake news persuades through heuristics, but real news convinces readers through arguments.

Textual analysis reveals the differences between legitimate articles and hoax articles on Wikipedia. Kumar et al. [20] investigate features to build machine-learning classifiers that can detect hoax articles, and compare the performance of the classifier with human rates. Human raters are told to identify hoax and non-hoax articles without getting any help from the web. Human accuracy is 66%, whereas the classifier achieves 86% accuracy on the same dataset, which indicates the significance of the features. Kumar et al. [20] identify four types of features: appearance features, link network features, support features, and editor features.

Textual patterns of satire articles are different from legitimate articles. Rubin et al. [21] examine the textual features of satire in four domains (civics, science, business, and soft news) and propose a SVM-based algorithm with five predictive features: Absurdity, Humor, Grammar, Negative Affect, and Punctuation. Absurdity feature

refers to the unexpected introduction of news in the final sentence. An article is considered humorous when the first and the last sentences of the article are not related to each other with respect to the remaining sentences. LIWC 2015 dictionaries are used to calculate the Grammar, Negative Affect, and Punctuation features. Among the 5 features, Absurdity, Grammar, and Punctuation feature combination perform well for detecting satire articles. Rubin et al. [21] also discover that punctuation marks and the individual textual features of shallow syntax (parts of speech) are highly indicative of the presence of satire.

Classifiers for detecting fake news perform well when the linguistic features rely on semantic information. Perez-Rosas et al. [22] extract several linguistic features to train a linear SVM classifier, and show that the classifiers that rely on the semantic information encoded in the LIWC lexicon show consistently good performance across domains. The following linguistic features are used in the study: ngrams, punctuation, psycholinguistic features, readability, and syntax.

The accuracy of fake news detection algorithms increases when the models consider both text and meta-data. Wang et al. [23] introduces a new dataset LIAR for fake news detection, which contains 12.8K short statements from POLITIFACT.COM. POLITIFACT.COM editors evaluate each statement for its truthfulness that has six fine-grained labels for the truthfulness ratings: pants-fire, false, barely true, half-true, mostly-true, and true. Speaker's current job, home-state, historical counts of inaccurate statements are also included with the text as the meta-data. Wang et al. [23] design a hybrid convolutional neural network and show that fake news detection algorithms give good results if the models consider both meta-data and text. For example, when the convolutional neural network model (CNNs) considers metadata with the text, it performs best than other models like logistic regression classifier (LR), support vector machine classifier (SVM), or bi-directional long short-term memory networks model (Bi-LSTMs).

### 2.1.2 Graph-Based Detection

This section presents the detection algorithms that consider network properties for distinguishing fake news. These algorithms also use other properties to improve the accuracy of the model.

To study how different individuals embedded in social networks initiate rumors, Arif et al. [24] analyze the rumor network during a hostage crisis in Sydney, Australia. For understanding the nature of rumors, Arif et al. [24] focus on three rumor related perspectives: 1. the volume perspective, 2. the potential exposure perspective, and 3. the content production perspective.

These three rumor related perspectives lead to four classifications of rumor effects: 1. Giant effect 2. Fizzle effect 3. Snowball effect 4. Babble effect. First, the giant effect occurs when a rumor gets high exposure and high volume. Official media and news outlets are examples of this effect as they have lots of followers; their posts reach individuals directly and have large initial footprints in the information network. Those posts easily get reposted and make a giant rumor effect. Second, the fizzle effect captures the moments where there is high exposure and low volume. For example, when an individual with high followers tweets information that is not shared or reposted by others. Third, the snowball effect starts with low initial exposure and eventually gets noticed by the crowd and diffuses at a large scale. Finally, the babble effect occurs when the individual with low followers posts a tweet that reaches a limited group to start a rumor.

Network properties of fake news can provide insight into the propagation of fake news. Starbird [25] generates a graph representing how users in the alternative media cite different web domains and studies the role of alternative media on fake news propagation. Starbird [25] also investigates the background of the content creators and categorize them based on their account type (e.g., mainstream, alternative media), Narrative Stance (e.g., supporting, denying), political leaning, and primary theme of

the content. The graph analysis reveals that the alternative media cite each other, and sometimes cite mainstream media to challenge that information or to get support for their theories. Results indicate the convergence tendency within the alternative media by echoing the same stories that target to undermine trust in information, rather than spreading a specific ideology.

Beside the network properties, detection algorithms can consider other properties to improve the performance. Qazvinian et al. [26] incorporate content-based features and Twitter specific features with the network properties, and identify that user history can be a good indicator of rumors and for belief classification (whether users believe or refute the misinformation). Qazvinian et al. [26] first extract tweets related to rumor, then identify users who believe or refute the misinformation. The dataset has 10,417 manually annotated tweets that are analyzed with three categories of features: Content-based Features, Network-based features, and Twitter-specific memes, such as hashtags, URLs.

### 2.1.3 Model-Based Detection

Model-based detection approaches learn how the news propagates so that it can simulate similar types of propagation. By capturing the difference in propagation between fake news and true news, detection algorithms distinguish fake news.

Model-based detection adopts epidemiological models that study the information diffusion by dividing the total population into several status chambers. The models also generate individuals' probabilities of changing their status. The most popular epidemiological model is SIS: 'S' stands for 'susceptible', 'I' stands for 'infected'. For example, initially, an individual is in a 'susceptible' state, but when the individual gets infected, his status changes to 'infected'. An individual can transition back and forth between 'susceptible' and 'infected' states in the SIS model. On the other hand, SEIZ model has four status: S (susceptible), E (exposed), I (infected), Z (skeptic). In the context of Twitter, those compartments/ status can be viewed as follows: 1.

Susceptible (S): represents a user who has not heard about the news yet, 2. Infected (I): denotes a user who has tweeted about the news, 3. Skeptic (Z): is a user who has heard about the news but chooses not to tweet about it, 4. Exposed (E): represents a user who has received the news via a tweet but has taken some time, an exposure delay, before posting.

SEIZ has a major improvement over the SIS as it incorporates exposure delay. Jin et al. [27] adopt SEIZ model to represent the information diffusion on Twitter and show that the SEIZ model parameters can separate rumors from real news. The primary analysis of the Twitter dataset on politics, terrorism, entertainment, and crime topics shows an activity burst immediately after the news becomes public. In contrast, for rumor cases, there are some occasional spikes instead of an increased tweets volume. The news network structure is more complicated than the rumor network; it indicates users can obtain real news from many sources, while users get the rumor information only from limited origins.

In summary, textual patterns show the difference between legitimate content and fake content. Identifying useful textual features can increase the accuracy of the detection algorithms. The accuracy can also increase when the algorithms consider other features with text. The graph-based detection algorithms are another approach to the detection of fake news that relies on investigating the network structure. These algorithms can also include textual features as graph properties to improve detection performance. The dissertation research relies on fact-checking services to identify the credibility of particular information and develop prompts the consider users' interaction patterns to nudge the interactions towards credible information and away from unverified information.

## 2.2 Media Literacy

This research adopts the techniques from Media Literacy to develop nudging prompts that communicate with the users of social media platforms as the primary focus of

the media literacy approach is to educate people about fake news or real news so that individuals can recognize it by themselves. Lewandowsky et al. [3] draw the attention that the ways to represent the correct information to people are also important. If people initially presume a piece of information accurate, later corrections of that information usually do not withdraw the first effect of the message. People tend to rely on the initial message, at least partially, even if they later learn the information is false. This phenomenon is called the continued-influence effect. This effect makes the later corrections less effective. These corrections can also occur back-fire effects if those directly challenge one's viewpoints. Lewandowsky et al. [3] suggest two important properties to consider for presenting the correct information effectively:

- The information should not directly challenge one's worldviews.
- People should be informed about the reasons why misinformation has propagated. If possible, corrections should provide meaningful explanations of the relevant event.

### 2.2.1 Effective Information Communication

To identify the effective ways of communicating correct information on the climate change issue Van der Linden et al. [28] study three approaches: descriptive text, a pie chart, and metaphorical representations to inform individuals that: '97% of climate scientists have concluded that human-caused climate change is happening'. The study includes 4 conditions: 1. Control group (n=115): participants see no information, 2. Descriptive text (n=87): participants see correct information in a form of plain text, 3. Pie chart (n=102): participants see correct information in a form of pie chart, 4. metaphors (n=800): participants see correct information in the forms of metaphors.

The results of Van der Linden et al. [28] study suggest that correct information should be short, simple, easy to comprehend and remember. For each condition of the study, participants are asked the question: 'To the best of your knowledge, what

percentage of climate scientists have concluded that human-caused climate change is happening?'. The question is asked twice; before and after showing the correct information. The positive changes in participants' responses indicate the effectiveness of the communication technique. Van der Linden et al. [28] find that descriptive text and pie chart approaches are better than metaphorical representations. The results also show that individuals, regardless of their political beliefs, change their beliefs about climate change when they see the correct information.

### 2.2.2 Communication in the Presence of Misinformation

The effectiveness of correct information gets reduced when individuals see misinformation. Van der Linden et al. [29] study how public beliefs on the scientific consensus are affected because of the misinformation and whether inoculation treatment can be useful in this situation. Inoculation theory suggests that individuals will be more resistant against misinformation if they can be exposed to weakened misinformation ahead. For the study, Van der Linden et al. [29] initially warn participants: 'some politically motivated groups use misleading tactics to try to convince the public that there is a lot of disagreement among scientists'. There are two inoculation conditions in the study: 1. general inoculation condition, 2. detailed inoculation condition. In the general inoculation condition, Van der Linden et al. [29] reiterate the claim that there is no disagreement in the scientific community that humans are causing climate change. For the detailed inoculation condition, additional arguments are used to support the claim. Participants (N=2167) are allocated to six conditions in the within-subject experiment:1. Control group: participants see no information, 2. Consensus treatment (CT): participants see only the correct information, 3. Countermessage (CM): participants see only the misinformation, 4. Consensus-treatment followed by countermessage (CT | CM): participants see correct information, then see misinformation, 5. Consensus-treatment + general inoculation followed by countermessage (ln1 | CM): participants see correct information, then general inoculation

messages, and finally see the misinformation, 6. Consensus-treatment + detailed inoculation followed by countermessage (ln2| CM): participants see correct information, then detailed inoculation messages, and finally see the misinformation.

The results of Van der Linden et al. [29] study indicate the effectiveness of inoculation treatment against misinformation. To find the most influential misinformation, Van der Linden et al. [29]) ask 1000 participants to rank six misinformation based on familiarity and persuasiveness. Participants identify the most influential misinformation is a petition from a website that says, 'Over 31000 American scientists have signed a petition stating that there is no scientific evidence that the human release of carbon dioxide will, in the foreseeable future, cause catastrophic heating of the Earth's atmosphere'. Van der Linden et al. [29] use the petition as the misinformation and measure the effectiveness of six conditions. The results suggest that when participants view only the correct information, their estimations of the scientific consensus increase. But when participants view misinformation, it negates the positive effect of the correct information. In that situation, inoculation messages are effective in preserving the positive effect of the correct information that remains the same across the political spectrum.

Cook et al. [30] extend the inoculation treatment study [29] to investigate how inoculation messages are effective for two different types of misinformation: 1. that casts doubt implicitly, 2. that casts doubt explicitly. In order to sow doubt implicitly, Cook et al. [30] design misinformation in the form of 'false balance' media coverage. The false balance strategy shows misinformation as a news article that first presents the mainstream scientific views, then presents contrarian scientists rejecting the mainstream arguments, and finally proposes an alternative explanation. Cook et al. [30] adopt that strategy to prepare misinformation for the study. For the correct information, a text-only description of different studies is used that reports 97% scientific agreement on human-caused global warming. Cook et al. [30] use another

textual description as the inoculation message that explains how the tobacco industry uses false balance strategy to confuse the public by presenting fake debates. To test the effectiveness of the inoculation message, 714 participants are randomly allocated to one of five groups: 1. Control group (n = 142): participants see no information, 2. Misinformation (n = 145): participants see only 3. Consensus/Misinformation (n = 142): participants first see correct information, then misinformation, 4. Inoculation/Misinformation (n = 142): participants first see inoculation message, then misinformation, 5. Consensus/Inoculation/Misinformation (n = 143): participants first see correct information, followed by inoculation message, and finally misinformation.

To cast doubt explicitly Cook et al. [30] adopt the fake experts strategy, where non-experts cast doubt on the expert agreement. Cook et al. [30] use the same misinformation from [29] study, but apply a different inoculation message that explains the techniques of fake experts. Cook et al. [30] show participants a tobacco industry commercial that features ten thousand physicians endorsing a particular brand of cigarette. The message also indicates that the physicians convey the impression of expertise without having relevant expertise. Thus, participants become inoculate against the fake expert technique. For the study, 392 participants are randomly allocated to four experimental conditions: 1. Control group (n = 98): participants see no information, 2. Inoculation (n = 98): participants see inoculation message, 3. Misinformation (n = 99): participants see misinformation, 4. Inoculation + misinformation (n = 97): participants first see inoculation message, then misinformation.

In both of the studies, Cook et al. [30] use participants' support for the free market as the proxy of their political ideologies, and measure the effectiveness of different interventions by comparing treatment conditions with the controlled condition. Result shows explicit misinformation increases more polarization than implicit misinformation. The results also indicate that inoculation technique reduces the negative effect

of both explicit and implicit misinformation.

In summary, how accurate information is delivered to the users is more crucial than the message itself. The short, simple, and easy to comprehend information or reports are effective ways to present accurate information. The information should not also challenge someone's beliefs about the topics. But the effectiveness of communicating correct information with users gets hampered in the presence of fake news. Inoculation techniques have potential to minimize the effect as the methods reveal the strategy of fake news to the users. This research direction focuses on educating and presenting correct information to the individuals without considering the affordances and users experience of social platforms. The research aims to extend the attempt of media literacy approach by developing nudging prompts that incorporate affordance with the media literacy findings to design interaction focused interventions for the social media users.

## 2.3 Nudging Prompts and Platform-based Interventions

The nudging prompts, without sacrificing the freedom of choice, alert users and provide individuals reminders, suggestions, and recommendations to steer their behavior in particular directions [31, 32, 8]. The social platforms can use nudging prompts to communicate with users to deliver related information about platforms' affordance and content. This section describes how the existing nudging prompts draw users' attention to the accuracy of information and credibility of the source but do not address users' interaction tendencies to make interventions personalized to their usage preference.

### 2.3.1 Nudging Users to Assess News Credibility

Researchers develop browser extensions to study nudging prompts that nudge users to assess news credibility. Bhuiyan et al. [8] develop a google chrome extension, FeedReflect [8] that prompts users to assess information credibility and veracity of

the source and warn users about the potential of misinformation. FeedReflect is designed for Twitter platforms that use visual cues and tooltips to highlight Tweets from mainstream sources and dim Tweets that are posted by non-mainstream sources. To make users conscious about their news consumption process, FeedReflect nudges users into conscious assessment of the credibility of information by showing a question from the comment section of the pertaining Tweet to engage users' reflective thinking. Bhuiyan et al. [8] study the effectiveness of FeedReflect to facilitate users assess the information credibility and use a survey button with the visual cues that contains questions related to news source credibility. 16 university students are divided into two conditions (treatment and control) and use the system for 3 weeks. Results show the effectiveness of nudging as the treatment group scored the tweets from mainstream sources higher and non-mainstream sources lower compared to the control group. The qualitative analysis reveals that FeedReflect [8] encourages users to rethink and reread the news, to use external resources for news credibility, and to actively participate in content credibility assessment.

### 2.3.2 The Accuracy Nudging Intervention

The survey questionnaires are designed to study the effectiveness of intervention that nudge users' attention to assess information accuracy. Pennycook et al. [33] find that people are good at identifying accurate headlines from fake ones and wish to avoid sharing misinformation (i.e., unverified or false stories). But the social media context shifts users' attention from identifying accurate information to other factors, such as partisan alignment. The goal of the accuracy nudging intervention is to shift users' attention to the accuracy of the information so that they judge the content's veracity before sharing that with others. As the intervention of the study, Pennycook et al. [9] ask participants to judge the accuracy of some non-COVID-19 related headlines before starting their COVID-19 related news sharing task. By following the steps in the treatment condition, Pennycook et al. [9] induce participants to

think about the notion of accuracy when they decide to share the news. The news sharing task of [9] study asks participants to submit their decision about whether they like to share the 30 headlines (15 true headlines, 15 false headlines) headlines on social platforms. Participants are presented with the headlines accompanied by pictures in the format of Facebook posts and asked, 'Would you consider sharing this story online (for example, through Facebook or Twitter?)' (yes/no). This self-report sharing measure is developed based on the observation from [34] that indicates the headlines reported by participants as higher likelihood of sharing also received more shares on Twitter. The results show the effectiveness of accuracy nudging as the participants of treatment condition share more true headlines that false compared to the participants of control condition.

### 2.3.3 Platforms' Affordances for Questionable Content

Social media platforms, such as Facebook, Twitter, and Google, use nudging prompts to communicate with their users in the context of questionable content distributed on the platforms. These platforms work with third-party organizations and develop machine learning algorithms to identify fake news and reduce fake news distribution on news feed [35, 36]. But for some information, when the social platforms have not identified the veracity of content or have not decided whether to remove the content due to its controversy, the platforms use fact-check alerts and provide additional related articles from fact-checkers to inform users about the credibility of information and its source [35, 4, 5]. Facebook adds related articles, including information from third-party fact-checkers, next to the questionable news to provide more perspectives about the information [Figure 2.1] [4], and Twitter informs users about the harmful content [Figure 2.2] [5]. The purpose of these nudging prompts is to increase users' contextual awareness about the information so that users can make informed decisions about their information consuming and sharing choices.

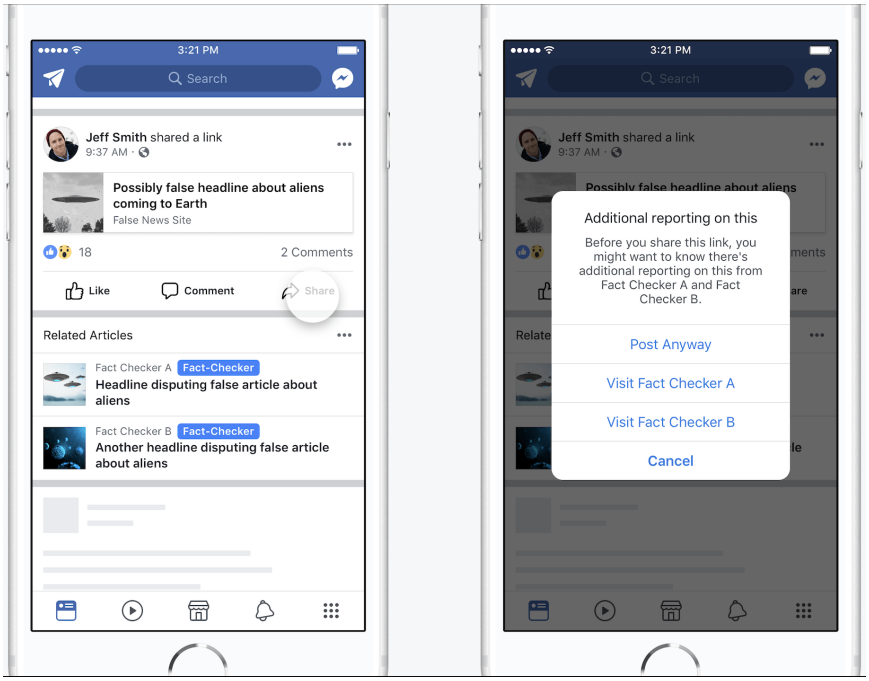


Figure 2.1: Facebook adds related articles with questionable content [4]

The prompts that aware users of the context of the questionable content reduce the distribution of the content. Facebook identifies the contextual awareness nudge that places related articles next to questionable news limits its distribution on the platform [4, 35]. Nekmat [7] has run an experiment to study the nudging effect of fact-check alert on sharing misinformation on social media. In a survey study, Nekmat [7] has divided 929 participants into 2 groups where participants in that group are exposed to the same information posted by either mainstream or non-mainstream sources. For each group, half of the participants are exposed to a fact-check alert prompt that says the information is disputed by fact-checkers. Results indicate that participants are less likely to share news posted by non-mainstream news than by mainstream news, but nudging prompts can also reduce the sharing of misinformation when the information is posted by mainstream news.

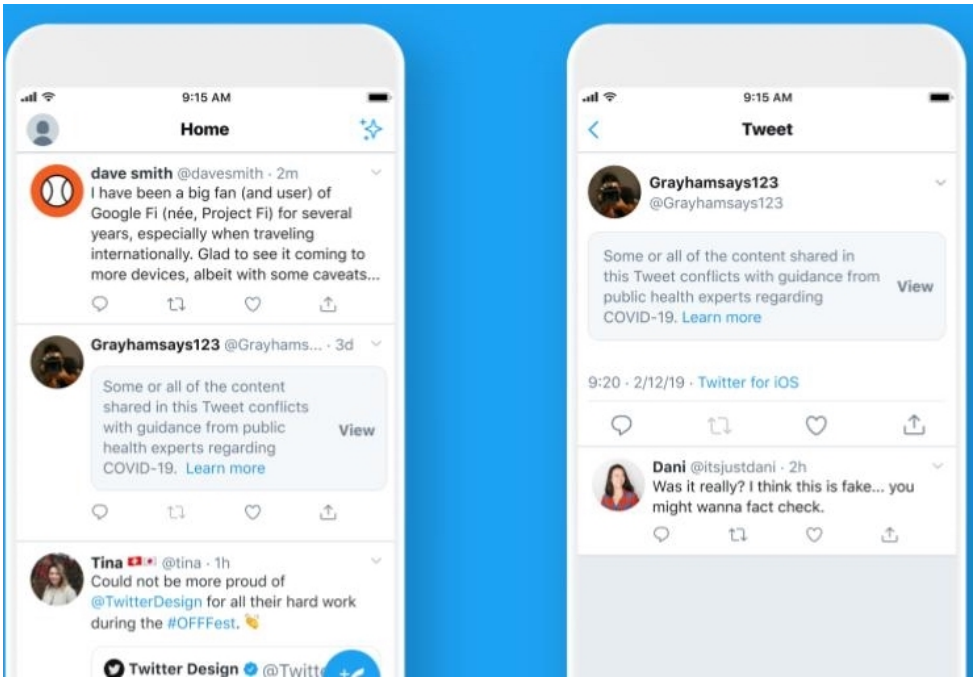


Figure 2.2: Twitter warns users about harmful posts [5].

In summary, the existing nudging prompts focus on drawing users' attention to the source's credibility and providing information to increase their contextual awareness. These prompts assist social media users in making informed decisions before consuming or sharing the news with other users. However, the prompts primarily focus on the factual position of information and do not assess users' preferred interaction habits with the content. This research considers users' interaction tendencies and content's factual positions to make interventions personalized to users' interaction tendencies.

# CHAPTER 3: INTERACTION FOCUSED INTERVENTION

Social media encourages users' participation which often leads to the spread of misinformation on social platforms. This chapter presents an Active-Passive (AP) framework that considers individuals' social media usage preferences when developing effective communication techniques to mitigate the viral spread of misinformation. This chapter describes the theoretical development of the AP framework, the association between the framework and users' online information-related behaviors, and shows the framework's implications in regulating nudging strategies for platform-based interventions.

## 3.1 The Active-Passive (AP) Framework

The Active-Passive (AP) framework has emerged from a review of the literature that describes users' social media interactions as a continuum from active to passive, where active users express high interaction tendencies compared to passive users. The framework has five interaction dimensions that connect the active-passive continuum with users' 3 types of social media usage: producing (e.g., users produce new content), participating (e.g., users share or rate content), and consuming (e.g., users read content), as illustrated in Figure 3.1. The AP framework distinguishes users and leverages their interaction tendencies to design usage-focused interventions and combat the spread of misinformation.

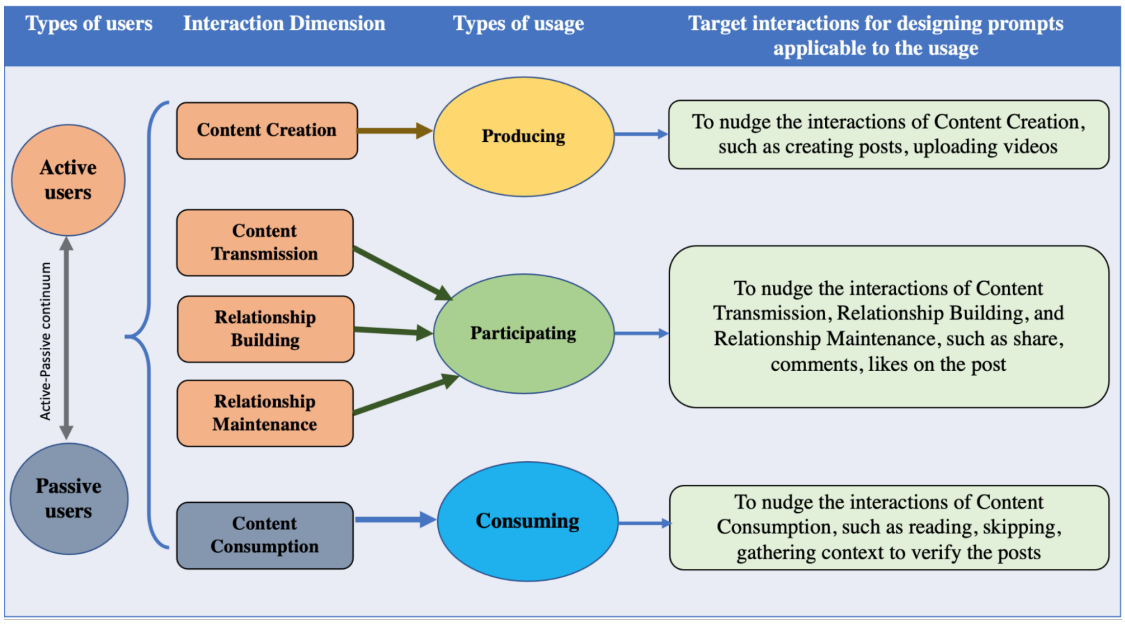


Figure 3.1: The AP framework addresses users' interaction tendencies to design usage-focused communication prompts [2].

The AP framework represents social media users with interaction habits ranging from active to passive. Active users produce original content, share information on social platforms, and maintain virtual relationships with communities and other users [11, 13]. Conversely, passive users consume content and avoid online interactions or participation with content and other users [13]. The framework has 5 dimensions of users' interactions: Content Creation, Content Transmission, Relationship Building, Relationship Maintenance, and Content Consumption. The first 4 dimensions of interactions are collected from [11], where Chen et al. [11] studied the active users' interactions on social platforms. One additional dimension of interaction, Content Consumption, is collected from [14], which studied users' interactions with misinformation. The framework uses the interaction dimensions to distinguish users based on their interaction patterns and to design prompts that can direct users' interactions to mitigate the spread of misinformation.

A user indicates the tendency of being an active user when their interactions on the platform spread across the 5 dimensions of interactions. Interaction items that

are used to create content, such as posting blogs, articles, photos, or videos, are in the dimension of Content Creation. Content Transmission includes interaction items such as sharing friends' posts/videos that are used to spread content on the platform. The Interaction items used to maintain online relationships, such as commenting on posts, chatting with friends through the platforms, are in Relationship Maintenance. Similarly, interaction items used to build virtual relationships, such as creating groups, sending invitations to friends and non-friend to join groups, are in the Relationship Building dimension. Finally, the content consumption includes users' interaction items associated with consuming content, such as scrolling, reading post/articles, and watching videos.

A user displays the tendency of being a passive user when most of their interactions are involved in the dimension of Content Consumption. Social media users are involved in social platforms in 3 ways: Producing, Participating, and Consuming [37]. The AP framework connects these three social media usage with the interaction functionalities associated with five interaction dimensions, as illustrated in Figure 3.2. The interaction items of the Content Creation dimensions described in [11] are associated with Producing usage. Content Transmission, Relationship Building, and Relationship Maintenance described in [11] are associated with Participating usage. Consuming usage is related to the content consumption dimension. Though these 3 kinds of usage can overlap while individuals interact with different types of content on social platforms, passive users have a strong preference for interactions towards consuming information.

The AP framework informs researchers to address Producing usage to direct users' interactions related to content creation, Participating usage to direct users' interactions associated with content transmission, relationship building, and relationship maintenance. Likewise, the AP framework enables researchers to focus on Consuming usage to develop communication strategies and nudging techniques while users are

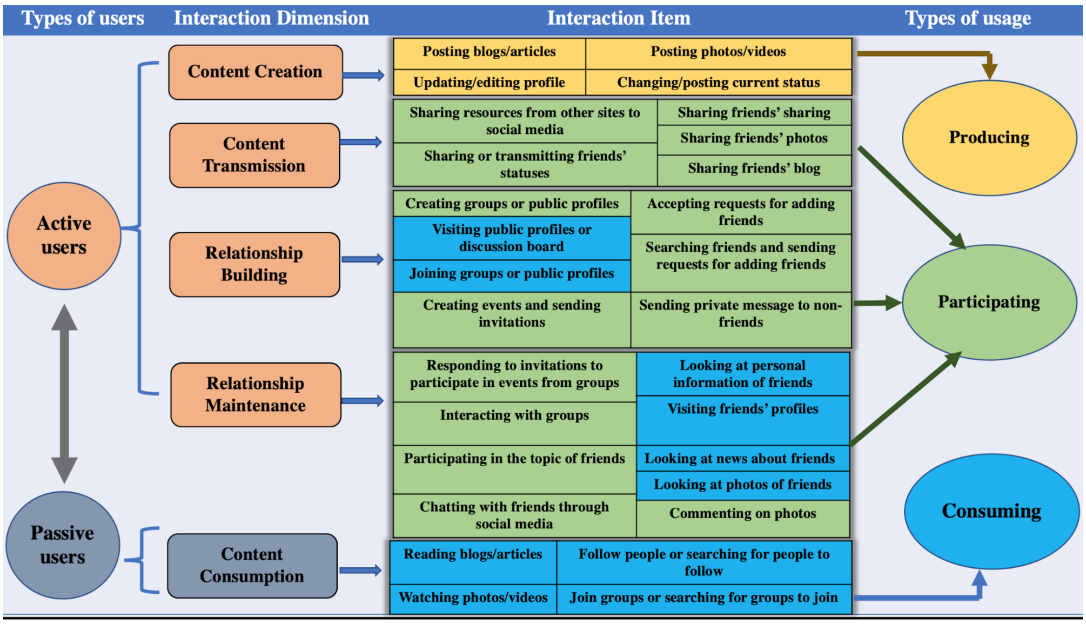


Figure 3.2: The AP framework identifies interaction functionalities associated with three social media usage to design interaction-focused interventions.

involved in consuming content. The framework provides a basis for researchers to design usage-focused prompts to communicate to users based on individuals' social media usage preferences.

## 3.2 Users' Interactions and Behaviors on Social Platforms

Users' interactions on social platforms construct the online behaviors exhibited on social media. The AP framework connects users' social media interactions and usage with their information-related behaviors so that the nudging prompts can direct the interactions to form healthier online behaviors and mitigate the negative effect of misinformation, as illustrated in Figure 3.3. This section describes the four information-related behaviors from reviewing the literature and shows the association between the framework and the 4 behaviors: information producing, sharing, seeking, and verification.

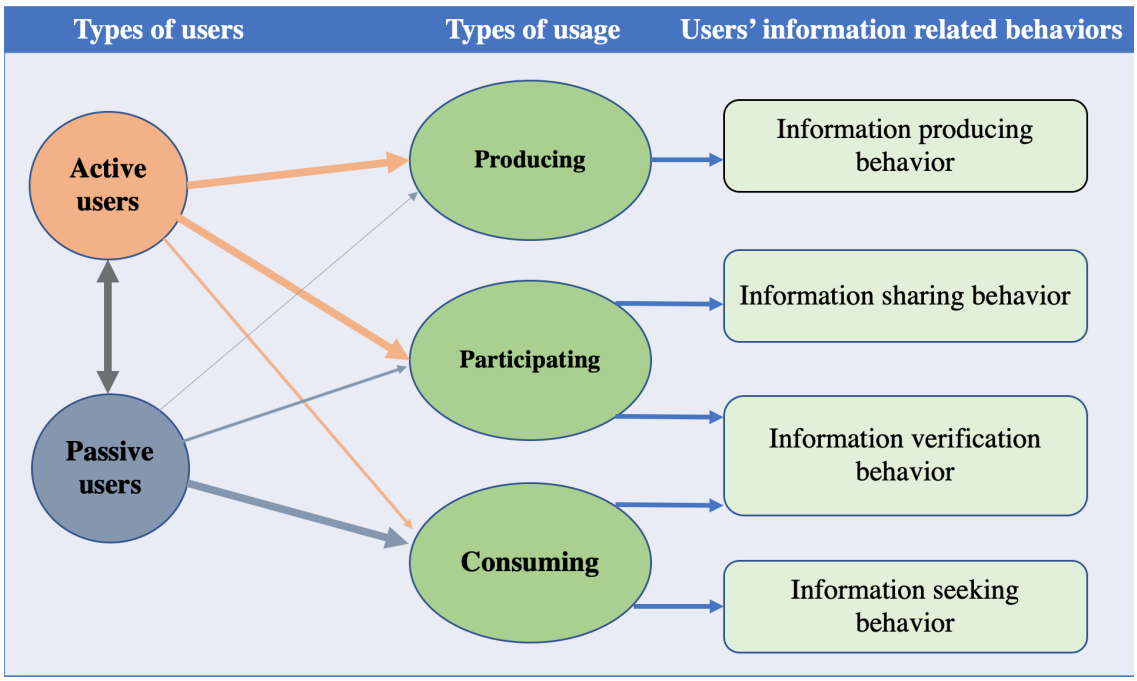


Figure 3.3: The AP framework connects users and usage patterns that create social media behaviors to design interventions for transforming users' behaviors.

### 3.2.1 Information Producing Behavior

Information producing behavior is associated with Producing usage of the AP framework. Users use the interaction functionalities related to the content creation dimension to create original content, such as posting articles or status and uploading photos or videos. Users produce content to satisfy their self-expression and self-actualization needs [37]. Self-expression refers to people presenting their inner-self to others and using the process to control others' impressions of them. Self-actualization means developing the expression that triggers behavioral goals, such as seeking recognition, fame, or personal efficacy.

Producing fake news on social media can result from self-expression and self-actualization needs. Wardle [38] refers to the inadvertent sharing of fake news as misinformation and the deliberate creation and sharing of fake news as disinformation. Tandoc Jr et al. [39] propose a topology of fake news definitions based on the content creator's intention and the level of the content's facticity and identify the

term fake news with six definitions: 1. news satire, 2. news parody, 3. fabrication, 4. manipulation, 5. advertising, and 6. propaganda. Immediate intention refers to the degree to which the content creator intends to mislead. For example, the creator of news satires and parodies intends to humor readers through some level of bending facts. These types of content assume an open disclaimer that the content is inaccurate. Conversely, the creator tries to mislead readers with fabricated and manipulated intentions. Without any disclaimer, the content creators deceive readers that the fake news they see is accurate, and their ultimate goal is to misinform people or attract clicks for advertising money.

### 3.2.2 Information Sharing Behavior

Information sharing behavior is associated with Participating usage of the AP framework. Users use the interaction functionalities related to the content transmission, relationship building, and maintenance interaction dimensions to share information on social platforms, such as sharing materials from other websites on social media and sharing friends' content.

The behavioral act of sharing information is usually assumed to be benefit-oriented [40]. Social exchange theory posits that people evaluate the cost and benefit before deciding to share information [41]. The cost could be using resources to accomplish sharing, such as the time devoted to interaction and the cues to understand the information's credibility. Benefits could be extrinsic or intrinsic [42]. Extrinsic benefits are related to social capital and reputation. For example, how users provide information on Twitter improves their reputation on social media. Intrinsic benefits could be the satisfaction originating from confirming one's ability to provide helpful information in a social network, such as providing feedback on products or movies.

People develop a worldview because of their social position and deep beliefs, influencing why people share fake news [43]. Media affect how users develop their worldviews, influencing how they extract meaning from the message written to be

interpreted in multiple ways. In addition, platforms' affordance influences users' information sharing behavior, and users prefer social networking and micro-blogging site platforms to draw other users' attention to their message [43, 44].

Users' retweet behavior on Twitter can be predicted based on the Heuristic-Systematic model (HSM) [45]. HSM suggests that individuals can process a message in two ways: 1. heuristically, which refers to processing messages quickly without involving heavy cognitive load, 2. systematically, which refers to processing messages carefully and deliberately. Liu et al. [46] use the HSM model and reveal that source trustworthiness, source expertise, source attractiveness, and the number of multimedia have significant effects on information retweeting. This result suggests that active users can cultivate these attributes to gain influence on social platforms, and both active and passive users can use these details as a cue for verifying information.

The uses and gratifications (U&G) approach is applied to study students' reasons for sharing misinformation that shows four main motivations [47]: 1. Entertainment: refers to using social media for personal enjoyment, 2. Socializing: refers to relationship development and maintenance with one's network on social media, 3. Information seeking: focuses on satisfying the need for consuming information, and 4. Self-expression and status-seeking: refer to expressing oneself and gaining a reputation. Self-expression and socializing are the primary motivations for students' sharing misinformation. Other top reasons are related to information characteristics, such as students sharing misinformation because the information could be a good topic of conversation, or the information looks exciting and eye-catching. In addition, students often share misinformation as they care less about the source's authoritativeness or the accuracy of the information.

### 3.2.3 Information Seeking Behavior

Information seeking behavior is associated with Consuming usage of the AP framework. Users use the interaction functionalities related to the content consumption

interaction dimension for reading news, posts, articles, or watching videos.

Users' online news seeking behaviors are shifting because of the social media platforms. Though the traditional way of seeking news online is to browse a news site or portal, this seeking behavior has been reduced to 34% [48]. Bentley et al. [48] investigate 174 participants' web browser logs to study Americans' online news seeking behaviors, and find two-thirds of the time users are exposed to stories by a link found in another site or shared to them by a friend. Participants normally receive biased information when the article comes from social media sites. Bentley et al. [48] reveal the polarization or filter bubble effect in users' news seeking habits as 48% of participants receive news mostly from left-biased sources, while 5% obtain it from right-biased sources. The analysis also shows that 47% of participants read the story from both sources, suggesting the AP framework's implication for understanding individuals' distinct information-seeking behaviors in 'consuming' usage.

### 3.2.4 Information Verification Behavior

Information verification behavior is related to how individuals verify information on social media. Individuals tend to verify information when they seek as well as share information [44], thus remaining involved in participating and consuming usage of the AP framework.

Individuals verify misinformation on social media in 3 ways [10]: First, participants rely entirely on the source's authority to judge whether the story is true. Second, participants formulate their beliefs based on their assessment of the news story and later utilize the source to support the belief. Third, participants purely rely on their assessment without considering the source. Participants often lack confidence in their ability to detect fake news and hesitate to trust the fact-checking tools to determine whether the information is fake or real [10].

Humans' information verification behavior model on social media proposed by Torres et al. [49] considers the nature of information, the epistemology of testimony, and

interpersonal trust in the network. When a user's social tie variation increases, the likelihood of exposure to fake news and the opposite of fake news also increases. Such disagreements lead the user into deception; increase an individual's awareness of fake news; decrease the trust in the network and media credibility. These attributes make individuals skeptical about the information and thus influence them to verify it on social platforms. Conversely, perceived cognitive homogeneity develops when individuals have a less diverse social tie and builds trust in the network, often influencing users to avoid the complexity of news validity. Consequently, individuals' awareness of fake news decreases, and they are less inclined to verify the information in their networks.

## 3.3 Implication of the AP Framework

The AP framework shows that the usage-focused communication prompts can be applied on 3 types of social media usage to direct the interactions associated with the usage. Fake news is a consequence of the type of usage in the AP framework called Producing. The spread of fake news is amplified by the type of usage in the AP framework called Participating when the user shares fake content. Likewise, users are trapped into filter-bubble and echo-chamber when they are involved in the Consuming usage of the framework. In this section, we describe the implications of the AP framework, as illustrated in Figure 3.4, in organizing the directions of the communication and nudging prompts to mitigate the negative effect of misinformation.

### 3.3.1 Nudging Directions for Consuming usages

The prompts for the Consuming usage focus on directing users' information consumption behavior so that individuals develop the habit of consuming verified information and remaining less affected when exposed to unverified information. Insincere consuming related interactions can be harmful to the users. For instance, users have a tendency to follow like-minded sources [50, 51, 52, 53], but these interactions can

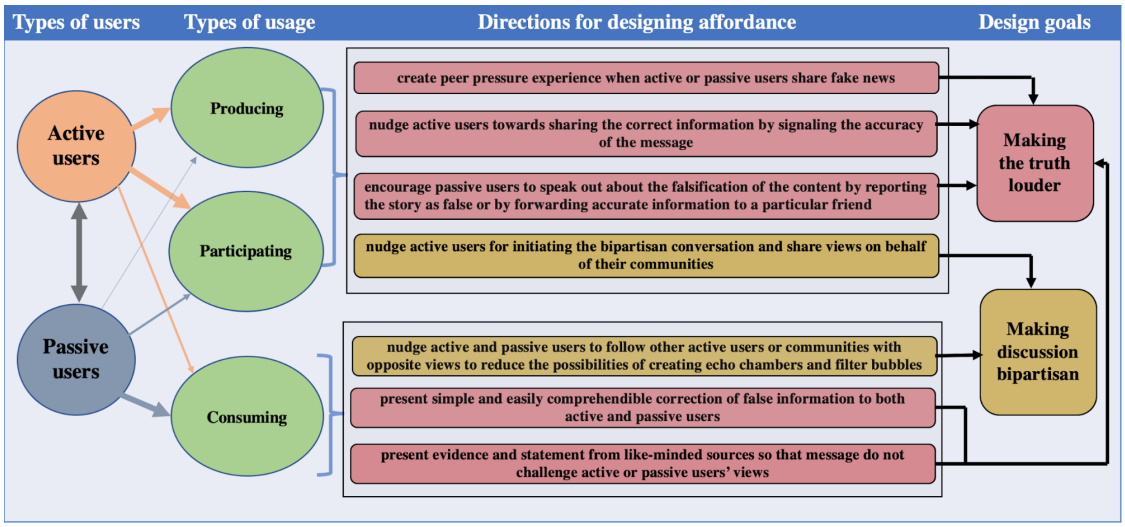


Figure 3.4: The AP framework leverages users' interaction tendencies for directing nudging prompts and designing usage-focused interventions to fight misinformation.

lead users to remain in echo chambers and be surrounded by filter bubbles. Nudging prompts that assist users in verifying information, getting additional context, and different perspectives are applied in the Consuming usage to nudge users' interactions in the content consumption dimension.

Interventions for Consuming usage can nudge users to follow the communities and users of opposite views. Instead of merely suggesting users follow others on social media, the interventions can communicate with users and inform them about the necessity of getting different perspectives on the topics, and how the interventions can help individuals to gain that perspective. The platform-based affordances that help users to verify information or get additional context are applied to the Consuming usage. The information button ('i') and the 'Related Article' section of Facebook [54, 55, 4], are applied when users are involved in the interactions related to the content consumption. Twitter warns users about harmful information when users are involved in Consuming usage [5]. The additional resources to support users in verifying information are also applicable for this usage.

The inoculation techniques [30, 29] that inoculate users against misinformation can be implemented as the interventions for Consuming usage. Similarly, presenting the

corrections of misinformation is also associated with the interventions of Consuming usage. As the corrections should not directly challenge individuals' worldviews [3] and people tend to accept the correction when the confirmation comes from similar ideological sources [52], interventions can present the statement and evidence about the corrections from the like-minded sources and active users. Curiosity has been applied as the basis of encouraging learners to learn [56] and can be the basis of platform-based interventions for Consuming usage and encourage people to be curious about learning the corrections and opposite viewpoints. Interventions of Consuming usage must limit the exposure of misinformation and should not show the misinformation in the context of correcting the message as the repeated exposure of misinformation can be harmful to the users [57].

### 3.3.2 Nudging Directions for Producing and Participating usages

Fake content is created and spread on social platforms when users are involved in Producing and Participating usage. Platforms such as Facebook and Twitter block social media accounts that misuse platforms' Producing and Participating usages to spread unverified information. In general, people prefer to share accurate information [9, 58] and interventions for Participating usage can highlight accurate messages to nudge the active users to contribute to the distribution of credible information. Bhuiyan et al. [8] developed a browser extension, FeedReflect, that uses visual cues to indicate whether the information source is mainstream or non-mainstream, and nudges users towards critical thinking before consuming the information. While FeedReflect [8] focuses on nudging users' information consumption behavior, similar kinds of approaches can be developed to direct users to distribute credible information and limit their interaction with unveiled information. Siddiqui et al. [1] developed 3 design principles to increase users' participation in spreading credible information and reduce users' participation in unverified content. Such interaction principles are applicable to the platform's Participating usage.

Insincere interactions of the Participating usage can lead to the spread of misinformation - active users due to their interaction tendency may often use the sharing functionalities in the context of unverified information. Facebook alerts individuals when the users press buttons to share any questionable content - such nudges can remind users to be reflective about their interactions and minimizes the insincere spread of misinformation. Creating peer pressure is suitable for these usages to reduce users' tendency to post or share misinformation. As the communication bridges across cultures foster the production of more neutral and factual content [16], interventions can nudge the interactions of the active users to get involved in the bipartisan discussions, make comments, and share their views on behalf of their communities. Communications can be personalized to the passive users by simplifying the interactions for them that require limited digital footprints, such as sharing the correct information to a particular friend or reporting the story as fake.

The nudging prompts for the active users can be applied in Producing or Participating usage to direct users' interactions in the usage for distributing credible information and limiting the spread of unverified information. Likewise, prompts for the passive users can be applied in the Consuming usage and focus on inspiring passive users to get involved in Participating usage and share credible information with their friends. Accordingly, the AP framework opens possibilities and empowers platform-based interventions to design communication prompts personalized to users' interaction tendencies that direct users' interactions towards mitigating the negative effects of fake news on social platforms.

In summary, this chapter presents the Active-Passive (AP) framework that distinguishes users based on their interaction tendencies and enables researchers to design usage-focused communication techniques and prompts to direct user interactions. The AP framework addresses social media users' active-passive tendencies and leverages the tendencies to assist users in developing healthy interaction behaviors that can

make the truth louder and discussion bipartisan. Accordingly, the framework allows researchers to develop and study the effect of communication and nudging techniques personalized to individuals' social media usage preferences that can mitigate the negative effect of misinformation on social platforms.

# CHAPTER 4: INTERVENTIONS PERSONALISED TO USERS' INTERACTION TENDENCIES

This chapter presents a theoretical basis and three principles for designing interventions personalized to users' interaction tendencies to make the truth louder on social media. This chapter describes the theoretical basis of transforming users' interaction behaviors, which builds on the Fogg behavior model (FBM) and addresses users' active-passive interaction tendencies of the AP framework to promote interaction behaviors for the social media users essential to make the truth louder.

## 4.1 Fogg Behavior Model

DJ Fogg proposes the Fogg Behavior Model (FBM) [59] that suggests a change of behavior happens when an individual has the motivation and ability to change the behavior, and a prompt occurs to trigger the target behavior. Individuals' behavior changes when the prompts can assist individuals in overcoming the behavioral activation threshold. The Fogg Behavior Model [59] offers 3 types of prompts: Signal, Facilitator, and Spark, for the individuals of different levels of motivations and abilities. The purpose of signal is to remind individuals about the target behavior, facilitator simplifies the target behavior, and spark influences individuals' core motivations to perform the behavior. As illustrated in Figure 4.1, the Signal type of nudging prompt is designed for individuals with high motivation and high ability, the Facilitator focuses the individuals with high motivation but low ability, and the Spark is for the individuals who have low motivation but high ability to perform the target behavior. These nudging prompts' success is also a timely issue as individuals' motivation and ability can change over time.

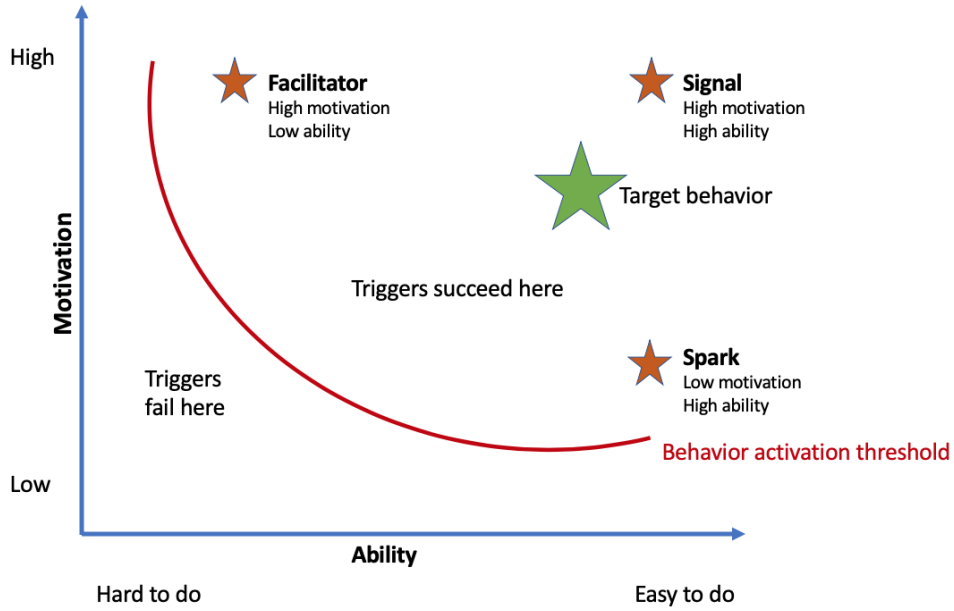


Figure 4.1: FBM offers three types of triggers for nudging users toward the target behaviors [59].

## 4.2 Nudging Directions to Make the Truth Louder

The Fogg Behavior Model inspires the 3 design principles to direct users' interactions across the active-passive continuum and assist users in adopting two target behaviors to make the truth louder on social platforms.

***Target behavior 1:*** users increase their interactions with verified information.

***Target behavior 2:*** users decrease their interactions with unverified information.

On social platforms, the ability to make the truth louder depends on users' interaction ability and preference on the platform. Active and passive users on social media demonstrate the opposite ability because of their online interaction tendencies. For example, the ability to contribute to the spread of verified information (target behavior 1) demands interactions with content and other social media users - such ability is high for active users but low for passive users, illustrated in Figure 4.2. Active users can interact with content more due to active users' natural inclination,

such as sharing information with other users, making comments, or sending love/like reactions to the content - these interactions contribute to the distribution of verified information. In contrast, passive users hesitate to perform such interactions and have low interactions with the content on social platforms, which makes adopting the target behavior 1 challenging for passive users.

Similarly, limiting the spread of unverified information (target behavior 2) is easier for passive users to adopt than active users, requiring users to interact less with the unverified content. For target behavior 2, passive users get an advantage as they generally tend to interact less with social media content, illustrated in Figure 4.3. In contrast, active users should be reflective about their activities so that they do not interact with the unverified content because of their natural behavioral tendencies, which makes adopting target behavior 2 harder for active users.

The effectiveness of a specific principle also depends on individuals' motivation to adopt the target behavior. To make the truth louder, the motivation of the target behaviors refers to the individual's motivation to combat fake news on social media and willingness to increase their interaction with verified information and reduce their interactions with unverified posts. High motivation indicates that individuals have the mental model to participate in combating misinformation, whereas low motivation indicates that individuals do not have the mental model to contribute and care less about the phenomenon of misinformation spreading on social platforms.

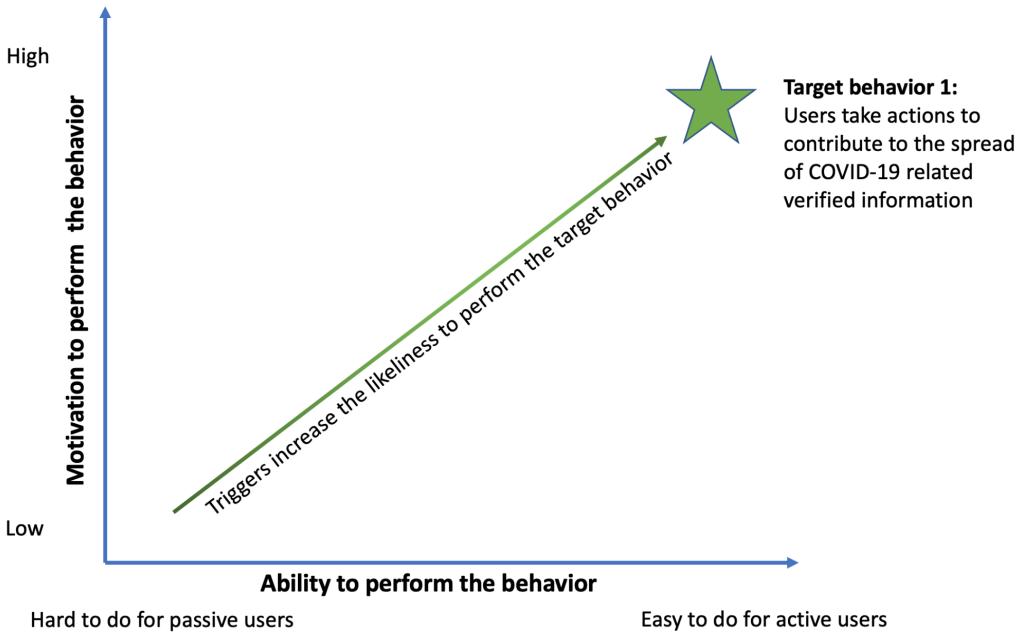


Figure 4.2: Target behavior 1 is easier for active users to perform but difficult for passive users [1].

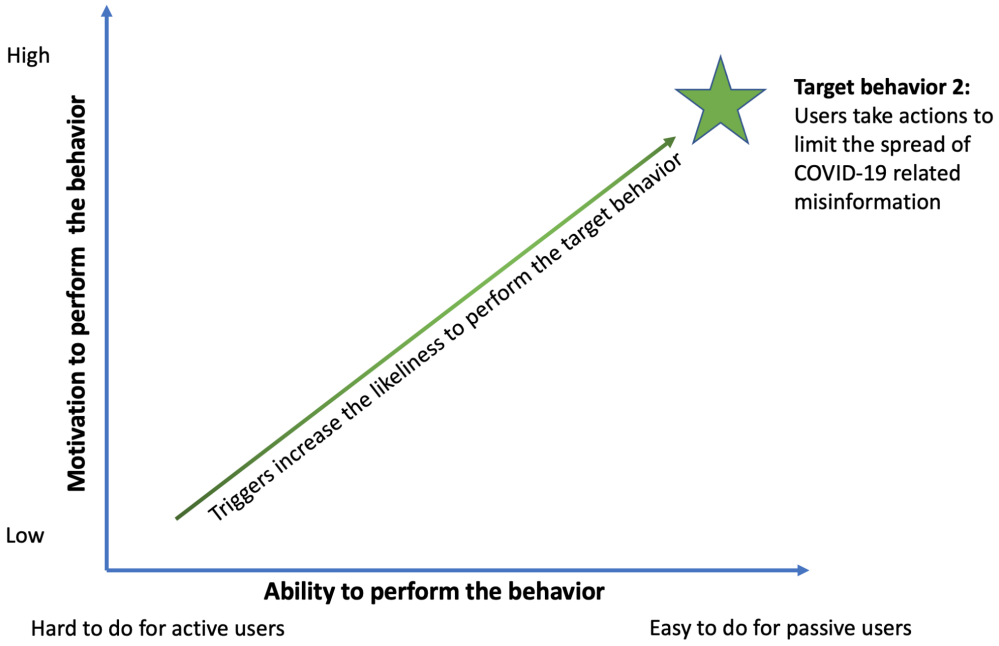


Figure 4.3: Target behavior 2 is challenging for active users to perform but easier for passive users [1].

Three categories of nudging prompt: 1. Contextual awareness on making the truth

louder, 2. Guidance on making the truth louder, and 3. Incentive on making the truth louder are designed to trigger social media users of various combinations of motivation and ability to achieve the target behaviors. We present and describe the purposes and differences between the prompt categories in the following section.

## 4.3 Design Principles Addressing Users' Interaction Tendencies

We present 3 design principles that address users' interaction tendencies: Awareness, Guidance, and Incentive on making the truth louder. In this section, we describe the design principles that appeal to the users of different levels of interaction abilities and motivation, and discuss the existing design interventions in reference to these principles.

### 4.3.1 Awareness on Making the Truth Louder

The purpose of the Awareness design principle is to assist social media users in recognizing verified and unverified content that appears in their social media feeds and remind users to perform the target behavior that can make the truth louder. This design principle is a Signal prompt in the FBM [59] that is effective for individuals who have high motivation and high ability to perform the target behavior. When active users on the platform have the motivation to participate in making the truth louder, they can respond positively to the intervention designs that follow the Awareness design principle promoting the target behavior 1. Likewise, passive users can respond easily to the interventions that follows the Awareness design principle to promote the target behavior 2 - requesting limited interactions with the unverified and questionable content.

Most of existing interventions can be described using the Awareness design principle as the focus of these interventions is to inform users about the context and validity of the information. For example, social media platforms, such as Facebook and Twitter, provide related fact-checkers' information so that users can get the context of

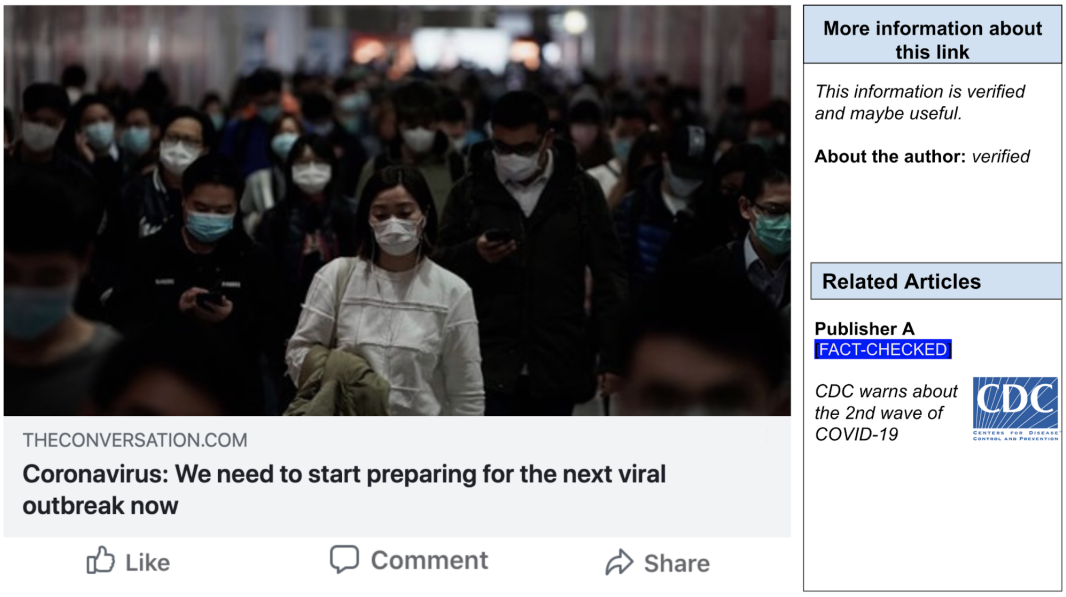


Figure 4.4: Prototype describing the Awareness design principle applied to verified posts [1].

the information. Facebook shows an indicator to related articles when the platform detects any questionable content [54, 4]. Twitter warns users if the platform identifies any harmful content [5]. The accuracy nudging intervention [9] draws users' attention to the accuracy of the content, and FeedReflect [8] applies visual cues to grab users' attention to the credibility of the information source - whether the information source is mainstream or non-mainstream. These platform-based interventions educate users about the context of the information when users are involved in content consumption interactions. Some interventions, such as Facebook, alert users when they interact to share any questionable content [4]. These interventions follow the Awareness design principle as the purpose is to make users aware of the context before they share the information on social media.

To describe the Awareness design principle, we present a design prototype that promotes the target behavior 1, illustrated in Figure 4.4. The prototype use the standard signifiers 'Like', 'Comment', and 'Share' buttons of Facebook that signal users can perform interactions to like the information, make comments about that


Figure 4.5: Prototype describing the Awareness design principle applied to unverified posts.

information, and share that information with other users. The credible information in Figure 4.4 is collected from [9] study, and we add 'More information about this link' and 'Related Articles' sections that assist users in getting the context of the content. Figure 4.4 follows the Awareness design principle that uses texts in the 'More information about this link' section to communicate with users about the credibility of information and information source, and have related fact-checked articles in 'Related Articles' section to provide more contextual information. This prototype focuses on informing the active users about context of the information so that the subset of active users who posses the motivation to contribute in making the truth louder become aware to share the verified information. Similarly, Figure 4.5 presents a prototype describing the awareness design principle applied on unverified posts.

### 4.3.2 Guidance on Making the Truth Louder

The purpose of the Guidance design principle is to simplify the interactions for the users to increase their ability to interact on social platforms and educate users about the interactions that can lead to the distribution of credible information and limit the

spread of unverified harmful information. This design principle is a Facilitator prompt in the FBM [59] that is effective for individuals who have high motivation but low ability to perform the target behaviors. This design principle focuses on promoting target behavior 1 among passive users by simplifying the interaction steps for them that assist their interactions for distributing credible information. Likewise, this design principle can promote target behavior 2 among the active users by designing interaction and affordance that assist them limiting their interactions with unverified content.

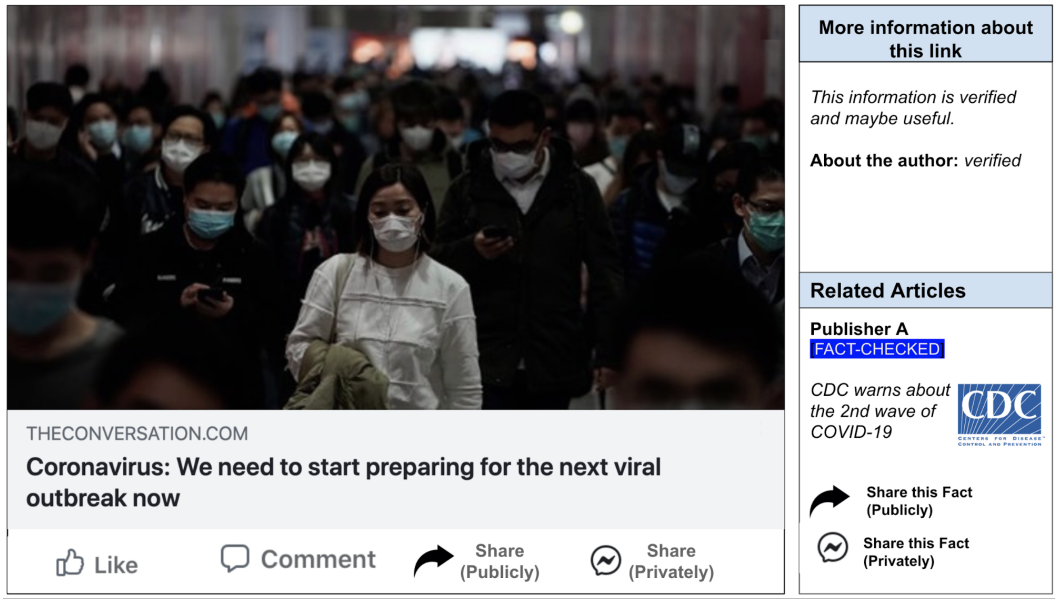


Figure 4.6: Prototype describing the Guidance design principle applied to verified posts [1].

To describe the Guidance design principle, we present a design prototype that promotes the target behavior 1 by simplifying sharing interactions, illustrated in Figure 4.6. The prototype follows the Guidance design principle that increases users' interaction ability with credible information by reducing the number of interaction steps required for sharing credible information. The Share button has the biggest impact on digital footprints as this functionality allows users to share the information with the users of their network; the Comments and Like buttons have smaller digital

footprints compared to that. Facebook includes different sharing options, such as share publicly or privately, and users get those sharing options when they press the share button.

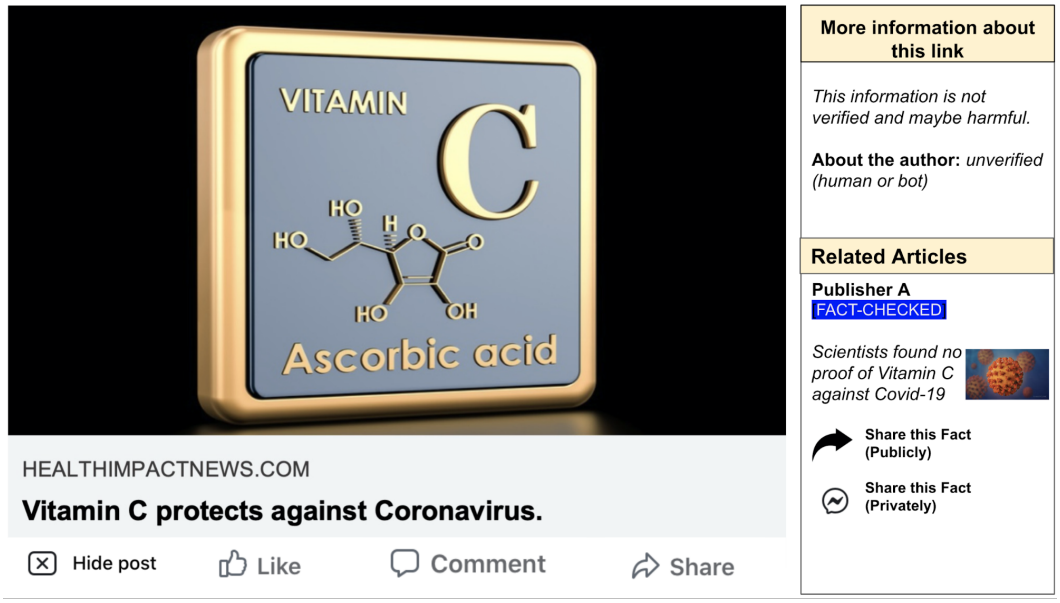


Figure 4.7: Prototype describing the Guidance design principle applied to unverified posts.

The prototype in Figure 4.6 presents different sharing options upfront to reduce the number of interaction steps for sharing. The prototype includes the privately sharing option to facilitate the motivated passive users' interactions toward the credible information. As passive users have a natural inclination to avoid digital footprint, the motivated passive users will feel comfortable sharing credible information privately to their friends rather than sharing publicly with the whole network. The prototype, Figure 4.6, includes additional 2 sharing options that enable users to share the verified fact-checked information with a single step of interaction. The design can apply visual cues on those buttons or use text to guide users about the interactions that lead to the distribution of credible information on the social platform. Similarly, Figure 4.7 presents a prototype describing the guidance principle applied on unverified posts that facilitates users' participation by adding several interaction functionalities, such

as hiding the unverified posts or sharing related fact-checked verified information.

### 4.3.3 Incentive on Making the Truth Louder

The purpose of Incentive design principle is to encourage and motivate users to orient their interaction behavior in a direction that can make the truth louder on the platform. The Incentive design principle is a Spark prompt in the FBM that is proven effective for the individuals who have the high ability but low motivation to perform the target behaviors. This design principle can prompt the less motivated active users to perform target behavior 1 and the less motivated passive users to perform target behavior 2.


Figure 4.8: Prototype describing the Incentive design principle applied to verified posts [1].

To describe the Incentive design principle, we present a design prototype that promotes the target behavior 1 by providing users badges, illustrated in Figure. 4.8. The concept of 'Community Service Badges' can demonstrate a way to incentivize social media users to increase their motivation for performing the target behaviors. When users perform social media interactions for making the truth louder, they will receive badges. The platform can add benefits to the badges, such as prioritizing the content

posted by users who have the badges, suggesting other users to follow the individuals who hold the badges due to the contribution in distributing credible information. Such platform-based benefits can attract active users to become reflective about their social media interactions and perform interactions only with credible information.


Figure 4.9: Prototype describing the Incentive design principle applied to unverified posts.

The platform-affordances can communicate with users and encourage them to participate in making the truth louder as a part of their responsibilities for creating personal, social, and societal impacts. Figure 4.8 includes the text *"Please participate in distributing credible information; your friends may benefit"* to communicate with social media users and inspire their motivation. As the community service badges indicate individuals' effort to make the truth louder on social platforms, the badges can gather positive impressions from other social media users, which can attract the platform's active users to attain the badges. The platform-based interventions can identify useful information and harmful information by relying on the fact-checking services and assist users in developing the interaction habits by rewarding them with the badges. Similarly, Figure 4.7 presents a prototype describing incentive design

principle applied on unverified posts.

## 4.4 Implication of the Design Principles

The design principles address the differences between active and passive users' online interaction tendencies to direct their interactions to make the truth louder on social media, as illustrated in Figure 4.10. Active users have high interaction abilities, and passive users have low interaction abilities. As users' interactions (e.g., shares, comments, likes) on social platforms lead to the distribution of the content in that platform, the principles intend to assist active users in adopting the target behavior to interact only with the credible information and not to interact with unverified or questionable information. Similarly, the principles intend to support passive users to increase their interactions with credible information that can make the truth louder on social platforms.

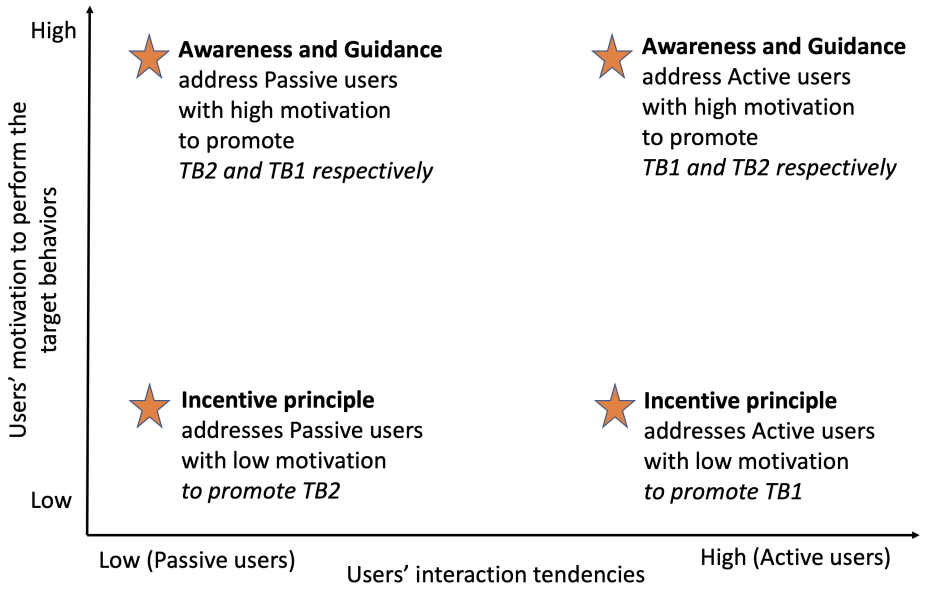


Figure 4.10: The design principles addresses the differences between active and passive users to perform the target behaviors [1].

The 3 design principles to make the truth louder that are adaptive to users' interaction tendencies. The relationships between the design principles and the credibility of the post, the target behaviors, and the user's interaction tendencies is shown in

Table 4.1: The design principles can measure the effectiveness of intervention designs for making the truth louder [1].

| **Factual status of the post** | **Target behavior to promote** | **User's interaction tendencies on the platform** | **User's motivation to contribute** | **Appropriate principle to apply on the post** |
|---|---|---|---|---|
| Verified posts | Target behavior 1 | High | Low | Incentive principle for active user |
| | | High | High | Awareness principle for active user |
| | | Low | High | Guidance principle for passive user |
| Unverified posts | Target behavior 2 | Low | Low | Incentive principle for passive user |
| | | Low | High | Awareness principle for passive user |
| | | High | High | Guidance principle for active user |

Table 4.1. These design principles encourage UX researchers to design and create affordances on the social media posts that are adaptive to individuals' interaction tendencies.

In summary, this chapter brings a shift in interpreting the problematic issue of misinformation spreading on social platforms as a design challenge for UX researchers to create platform-based affordances that encourage users to adopt new interaction behaviors: interact more with credible content and interact less with harmful content. Instead of solely relying on reducing the spread of misinformation, this research encourages UX researchers to explore design ideas of the three design principles by addressing the difference between active and passive users to create affordances that direct users' interactions to the truth louder on social media.

# CHAPTER 5: EXPERIMENT DESIGN TO STUDY THE PERSONALIZED INTERVENTIONS

This chapter explains the survey study designed to investigate the effect of 3 design principles on users with active-passive interaction tendencies. This chapter presents how we conducted the study and collected data. Finally, this chapter shows the analysis methods to find answers to the research questions of this thesis.

## 5.1 Study Design

This section describes the survey design conducted on Amazon Mechanical Turk from November 30 to December 12, 2021, which has 2 experiment conditions: control and treatment. The control condition adopts the awareness design principle, and the treatment condition adopts the guidance design principle. This section explains the designs used in 2 conditions and the questionnaire to gather insights into the incentive design principle. This section describes the study tasks for the 2 conditions, headlines selected for the study tasks, the participants' recruitment process, and the study's social media usage and demographic questionnaire.

### 5.1.1 Study Tasks for Awareness Design Principle

The awareness design principle aims to inform participants of the factual position of the post they are seeing. The design presenting the headline as a social media post adopts the awareness design principle discussed in Chapter 4. Figure 5.1 shows an example of the design used in the controlled condition when the headline is unverified. Participants can see the headline in plain text; we do not include the source of the headline or any image to reduce biases. To inform the participants about the factual position of the headline, we use the information icon "(i)" and the text message "This

information is determined as False by politifact.com". We added another section named' Related Articles' to give the participants additional information about the headline. The 'Related Articles' section includes the fact-check headline of an article and supports the label (False or True) determined by politifact.com [60]. A similar design pattern is used for the verified content illustrated in Figure 5.2. We have finalized the designs for the survey study by addressing feedback from the participant of the pilot studies in several iterative process.


Figure 5.1: Design and interaction functionalities of the awareness principle applied to unverified headline (control condition).

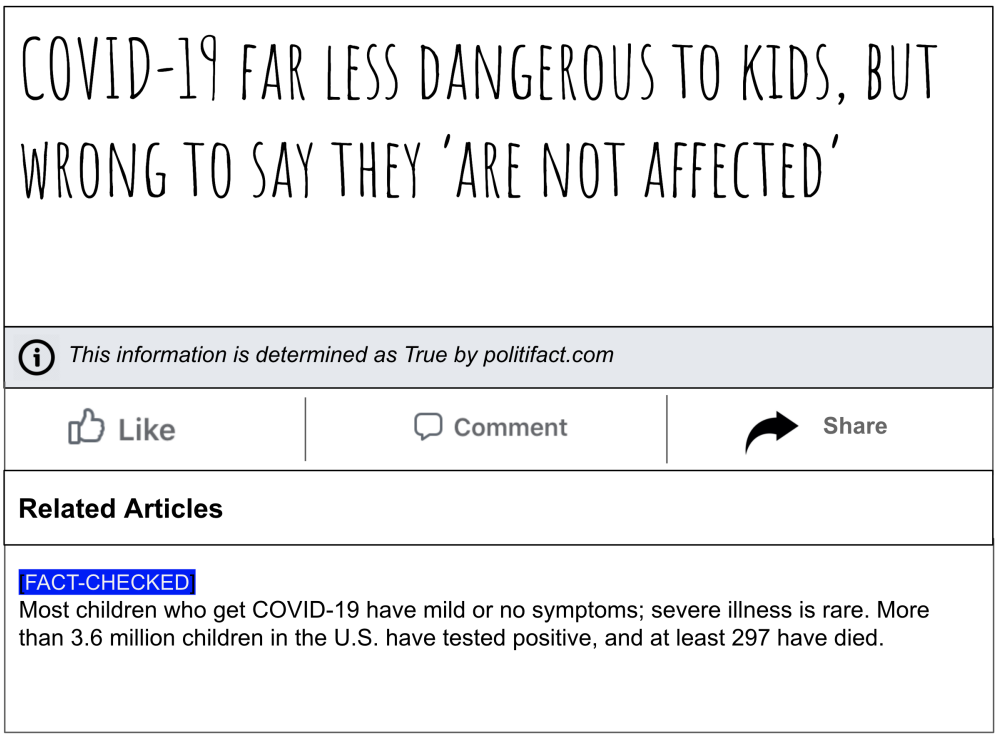


Figure 5.2: Design and interaction functionalities applied to verified headline in the awareness principle (control condition).

The design includes three basic interaction functionalities: Like, Comment, and Share - these functionalities are the most common interaction facilities provided by the platforms. The Like button allows users to react to the post, and users use the Comment button to express opinions about the post. The Share button allows users to spread the headline on the platform. After showing participants the headlines, we ask them the question: "What actions would you like to take?" and provide them with 4 multiple options: 1. Press 'Like' button, 2. Press 'Comment' button, 3. Press 'Share' button, and 4. Take no action. The study task is inspired by the news sharing task used in [6, 9, 7], where participants see news headlines and are asked to decide whether they would share the headline on social media or not. On the contrary, our study task asks participants to decide which interaction functionalities they would like to use if they see the headlines (i.e., post) on their social media.

Additionally, the survey includes a question to collect information regarding what

perceptions participants have toward the post: whether participants think the post conveys helpful information or harmful information. We allow participants to select both helpful and harmful options of the same headline - this study choice is informed by the insights from the pilot study. In the pilot studies, we observed that participants found the same headline as helpful and harmful, depending on the audience participant would like to communicate with. Participants also mentioned that the context often influences whether they would label the post as helpful or harmful. We also allow participants to choose 'Not Sure' option if they can not decide. Similarly, we collect information regarding whether participants find the posts interesting or relevant. People like to share content when they find the information interesting or relevant [6]. This study investigates whether individuals have interaction preferences when they find any post interesting or relevant.

### 5.1.2 Study Tasks for Guidance Design Principle

The guidance design principle aims to assist users in spreading credible information. The design that adopts the guidance principle for verified posts provides two post-sharing functionalities to share the verified posts, as illustrated in Figure 5.3. Users can share the verified posts publicly or privately. Additionally, the design provides two article-sharing functionalities that allow participants to share the related fact-checked information. Participants can share the fact-checked article publicly or privately. The observations during the pilot studies inform the design decisions of publicly and privately sharing functionalities - participants tend to utilize the publicly and privately sharing functionalities in different scenarios. Thus, the design includes these two functionalities to assist users in sharing more credible information.


Figure 5.3: Design and interaction functionalities applied to verified headline in the guidance principle (treatment condition).

The design for unverified posts that adopts the guidance principle provides two article-sharing functionalities to spread credible information related to the unverified posts. As illustrated in Figure 5.4, participants receive the interaction functionalities to share the fact-checked article publicly or privately. The design also keeps the traditional post-sharing functionality that users can utilize to share unverified posts. However, the post-sharing functionalities when the posts are unverified do not include the privately or publicly sharing functionalities and do not assist participants in sharing unverified posts. We have finalized the designs for the survey study by addressing feedback from the participant of the pilot studies in several iterative process.


Figure 5.4: Design and interaction functionalities of the guidance principle applied to unverified headline (treatment condition).

Before the study tasks begin in the treatment condition, the survey educates participants about the functionalities of the additional sharing buttons presented in the study. After showing the participants the headlines, the questionnaire asks: "What actions would you like to take?" and provides participants with multiple options to select interactions presented with the headlines. Participants receive options when posts are unverified: 1. Press 'Like' button, 2. Press 'Comment' button, 3. Press 'Share' button, 4. Press 'Share Article (Publicly)' Button to share Related Article, 5. Press 'Share Article (Privately)' Button to share Related Article, 6. Take no action. When posts are verified, participants see 2 post-sharing options: Press 'Share (Publicly)' Button to share social media post and Press 'Share (Privately)' Button to share social media post, instead of the Press 'Share' button option.

### 5.1.3 Questionnaire for Incentive Design Principle

The incentive design principle focuses on users' platform-based incentives and motivation that can increase participation in combating misinformation. This section describes the questionnaire designed to gather insights regarding users' preferences for the platform-based incentive across the active-passive continuum. The questionnaire includes questions regarding participants' motivation to participate and their level of trust in fact-checking journals and correlates the responses with the interaction decisions across the active-passive continuum. All these questions are presented as the study's post-task questionnaire.

Four kinds of platform-based incentives are identified during the rounds of pilot studies: 1. Getting badges, 2. Getting followers, 3. Content prioritization, and 4. Receive information regarding the impact. The observation during the pilot studies reveals that individuals might be encouraged to participate in combating misinformation if platforms offer them badges, help them gain followers, prioritize their content and inform them how their participation helps their community. Table 5.1 shows the statements used in the survey study so that the research can investigate users' preference for the platform-based incentives due to their active-passive interaction tendencies.

Table 5.1: The platform-based incentives and corresponding statements.

| **Platform-based incentives** | **Statements of the incentive** |
|---|---|
| Getting badges | "The platform gives me badges that inform other users about my contribution for combating misinformation." |
| Getting followers | "The platform suggests other users to follow my account as I contribute in combating misinformation." |
| Content prioritization | "The platform prioritizes my posts to other users as I contribute in combating misinformation." |
| Receive information regarding the impact | "The platform shows me how I am helping my friends and community by participating in combating misinformation." |
| Other | "Other" |

The questionnaire collects information regarding individuals' level of motivation to contribute to combating misinformation that requires participating in spreading credible information and reducing the spread of misinformation. In the post-task questionnaire, participants are asked to rate their level of agreement with two statements: 1. "I like to contribute to sharing verified and helpful information on social platform." and 2. "I like to contribute to reducing the spread of unverified and harmful information on social platform." For those two statements, participants can choose their agreement using one of the options: Disagree, Somewhat Disagree, Neither agree nor disagree, Somewhat Agree, Agree.

The questionnaire collects information regarding individuals' levels of trust in the fact-checking journals as the interventions present the factual positions determined

by these journals. In the post-task questionnaire, participants are asked to rate their agreement with the statement: "When judging the credibility of a news article, I trust information from the fact-checking journals (e.g., Politifact.com)." Participants are allowed to rate their level of agreement using the options: Disagree, Somewhat Disagree, Neither agree nor disagree, Somewhat Agree, Agree.

### 5.1.4 Pilot Studies

Several rounds of pilot studies inform the development of the survey study. The UI designs used in control and treatment conditions adopting the awareness and guidance design principles respectively get updated from the participants' feedback in the pilot studies. The questionnaire developed for the platform-based incentives is prepared from the responses in the pilot studies.

The pilot studies conduct interviews to get participants' feedback on the design and collect their responses to the survey questionnaire. Three rounds of pilot studies: 1st round included 6 participants, 2nd round included 4 participants, and 3rd round included 2 participants. Each round gets updated on the participants' feedback from the previous rounds, and we finalize the survey study in the 3rd round of the pilot study.

The UI aspect of the design principles is developed throughout an iterative process that addresses the feedback from the pilot study's participants. Participants in the pilot studies provide feedback on the design layout. For instance, participants' feedback was to put the 'Related Article' section underneath the social media post. Initially, the 'Related Article' section was on the right side of the post. We address participants' feedback and put the related article section underneath the social media post. Participants provide feedback on the labels of the UI buttons. We address the feedback and modify the labels of the UI buttons so that participants of the survey study can understand the functionality of the buttons after reading the labels. The survey study includes a tutorial to educate participants about the design layouts and

buttons' functionality.

Participants' responses during the pilot studies guide the development of a platform-based incentive questionnaire for the survey study. We identified the platform-based incentives that participants in the pilot study mentioned during the interviews, and these incentives were included in the incentive questionnaire designed for the survey study. As the pilot study does not have IRB approval, we do not report any participants' responses in this research.

### 5.1.5 Headlines Selected for the Study as Social Media Posts

This study focuses on Covid-19 vaccine topic and address the helpful and harmful aspects of the vaccine in regard to individuals' health and society. We carefully pick the headlines that signal minimum political tone so that the study has less political biases. Initially, we selected 60 verified and unverified headlines from politifact.com, a well-known 3rd party fact-checking journal that distinguishes true and false information posted on social media. The politifact.com journal [60] labels the posts as one of the 6 categories: True, Mostly True, Half True, Mostly False, False, and Pants on Fire. We rely on these labels to determine the factual position of the post.

For the study, we selected 8 health-focused headlines - 4 labeled as True and 4 labeled False by the politifact.com journal. We pick the recent headlines from the politifact.com website that are relevant to the covid-19 vaccine. The order in which participants see the headlines during the study is random. Each time when participants see the headlines, we prompt them with 2 questions that collect the action participants would take after seeing the information and their perception of the information. This study design was approved by the University IRB committee.

Table 5.2: Headlines that are selected for the study from the politifact.com journal.

| **Headline ID** | **Headline** | **Factual position** |
|---|---|---|
| T1 | San Francisco had twice as many drug overdose deaths as COVID deaths last year. | True |
| T2 | COVID-19 far less dangerous to kids, but wrong to say they 'are not affected' | True |
| T3 | For mitigating COVID-19 spread, 'masks (are) the number one way to do so.' | True |
| T4 | COVID-19 vaccines work, even if they aren't 100% effective | True |
| F1 | 15-year-old boy passes away from heart attack two days after Pfizer COVID-19 experimental jab. | False |
| F2 | The death rate for fully vaccinated people is significantly higher than non-vaccinated people. | False |
| F3 | If you get the COVID-19 vaccine, you can't donate blood or plasma 'because it's completely tainted.' | False |
| F4 | 60% of new COVID-19 patients are people who received the vaccine. | False |

### 5.1.6 Participants

We collected 1075 responses and randomly assigned participants into two design conditions: Condition A (control condition) was the baseline condition typical of most social media platforms, and Condition B was the treatment condition in which we encouraged interaction with verified news and discouraging interaction with unverified

news. The inclusion criteria were that participants should be 18 years of age or older, use social media at least 3 times a week, read Covid-19 Vaccine-related news, be located in the US. In the survey, we included 2 additional questions to check participants attention while completing the survey. We eliminate the survey records that have incorrect responses for these two attention-check questions. In addition, we remove the records that have not finished answering all the questions and have incomplete responses. Finally, we had 1006 participants in both conditions (N=503).

### 5.1.7 Social Media Usage Questionnaire

The survey includes 5 questions related to participants' social media usage tendencies. The self-reporting questions collect information on how individuals spend their time on social platform: whether they like to create content, spread content, or consume content. We collect information about how likely participants what to spend their time on building or maintaining relationships with other social media users. These 5 questions are developed from the literature review presented in Chapter 3 for describing the Active-Passive (AP) Framework. Table 5.3 shows the social media questionnaire used in this study. Unlike the previous studies [6, 7, 9] that collect certain information regarding users' social media usage, this study is designed to utilize the social media usage questionnaire to determine users' interaction tendencies in the active-passive continuum of the AP Framework.

### 5.1.8 Demographic Questionanire

The survey collects participants' demographic information regarding their gender, age, the social media platfomrs they use. Participants can choose multiple options from the choices such as Facebook, Twitter, Snaptchat, Instagram, and are given the option to sumit the names of other platforms that are not listed as the options. We collect information about how proactiveily partciipants search for the Covid-19 veccine related information with the options: Never, Rarely, Sometimes, Often, Very

Table 5.3: Social media usage questionnaire to determine users' active-passive tendencies.

| **Interaction dimension** | **Corresponding statement for the dimension** |
|---|---|
| Content creation | On social media, I spend time creating my own content (e.g., posting tweets, status, articles, photos, videos). |
| Content transmission | On social media, I often share information (e.g., retweet, share content created by others). |
| Relationship building | On social media, I spend time on relationship building (e.g., create group/event, join group/event, sending private messages to non-friend). |
| Relationship maintenance | On social media, I spend time on relationship maintenance (e.g., commenting on content, chatting with friends). |
| Content consumption | On social media, I spend time browsing content created by others. |

Often. We also collected participants' political position, though the research does not study any political biasis.

## 5.2 Data Collected

This section presents the data collected from participants and their responses to the social media usage questionnaire. This section shows the internal consistency and the findings of exploratory factor analysis on participants' responses to the social media usage questionnaire. This section presents the distribution of participants' interactions and perceptions of the verified and unverified headlines.

### 5.2.1 Participants' Demography

In the control condition, there are 503 participants. Among those participants, 286 are male, 211 are female, and 6 individuals prefer not to provide answers to the gender question. The mean age distribution of the 503 participants is 38 and the standard deviation is 11. The minimum age is 20 and the maximum age is 71.

In the treatment condition, there are 503 participants. Among those participants, 305 are male, 195 are female, and 3 individuals prefer not to provide answers to the gender question. The mean age distribution of the 503 participants is 38 and the standard deviation is 11. The minimum age is 20 and the maximum age is 89.

In the control condition, most of the participants use Facebook (88%) and Instagram (75%). After those 2 platforms, the participants mention using Twitter (71%) and Snapchat (28%). Finally, We notice that the participants mention using other social media accounts 7% time, which indicates that popular social media platforms are listed as the survey options. Participants in the treatment condition have the similar distribution.

Table 5.4: Distribution of the social media platforms used by the participants.

| **Social media platform** | **Participants (control condition)** | **Participants (treatment condition)** |
|---|---|---|
| Facebook | 441 | 439 |
| Instagram | 375 | 388 |
| Twitter | 355 | 347 |
| Snapchat | 143 | 184 |
| Other | 37 | 36 |

Majority of the participants in the controlled condition state that they proactively search for Covid-19 vaccine-related information. 20% of the participants search for the information very often, 35% of them are often, and 31% of the participants mention that they sometimes proactively search for the information. The political position of the participant sample is 57% democrat, 26% Republican, and 17% of the participants are Independent.

Most of the participants trust the factual position of the post determined by the fact-checking journals such as politifact.com. 35% of participants agree with the statement that they trust the fact-checking journals and 44% of participants respond that they somewhat agree with the statement. In contrary, we find 6% of the participants disagree with the statement and 4% of the participants state that they somewhat

agree with the statement.

### 5.2.2 Internal Consistency of Social Media Usage Questionnaire

Internal consistency evaluates the reliability of a questionnaire, and this study calculates the internal consistency using Cronbach's $\alpha$ coefficient that computes the correlations between all pairs of question items [61, 62]. Cronbach's $\alpha$ of social media usage questionnaire is 0.83 when 4 dimensions are considered: Content Creation, Content Transmission, Relationship Building, and Relationship Maintenance. Cronbach's $\alpha$ becomes 0.76 when the Content Consumption dimension is considered in addition to the 4 dimensions. As participants' across the active-passive continuum use the content consumption dimension, the $\alpha$ score gets reduced when that interaction dimension is included in the calculation. However, both of Cronbach's $\alpha$ scores are in the acceptable range [62]. In general, the reliability of the questionnaire is acceptable when the $\alpha$ is between $0.6 - 0.8$ and good when the value is greater than 0.8 [62]. However, the higher $\alpha$ values (e.g., 0.95) are not necessarily good as these values indicate redundancy in the questionnaire [62, 63]. This study considers the 5 dimensions of interactions instead of 4 to capture users' active-passive interaction tendencies on the platform.

### 5.2.3 Exploratory Factor Analysis on Social Media Usage Responses

The factor analysis is applied to explore the factors emerged from participants' responses to the social media usage questionnaire and find 3 factors. The 1st factor have top 2 weights for relationship maintenance and relationship building. This factor also has larger weights for Content Creation and Content Transformation dimensions. The 2nd factor has high weights only for the content creation and transformation dimensions - weights in other dimensions are low for the 2nd factor. Finally, the 3rd factor has the highest weight for the content consumption dimension and the lowest weight for content creation dimension.

Table 5.5: Thee factors have emerged from participants social media usage responses.

| | **Factor 1** | **Factor 2** | **Factor 3** |
|---|---|---|---|
| Content Consumption | 0.0338411 | 0.01482546 | 0.40107748 |
| Content Creation | 0.58185 | 0.53013142 | -0.16346337 |
| Content Transmission | 0.47518341 | 0.58886818 | 0.16743347 |
| Relationship Maintenance | 0.66556763 | 0.22931341 | 0.16707236 |
| Relationship Building | 0.72888644 | 0.35448521 | 0.03759299 |

### 5.2.4 Relationships Between the Interaction Dimensions

The 4 interaction dimensions: Content Creation, Content Transmission, Relationship Maintanence and Relationship Building are positively correlated with each other. In contrast, the content consumption dimension does not have any significant relationship with the 4 interaction dimensions. For example, content creation is positively correlated with content transmission (0.56, $p < .001$), Relationship Maintenance (0.48, $p < .001$), and Relationship Building (0.60, $p < .001$). Similarly, Content Transmission dimension is positively correlated with Relationship Maintenance (0.47, $p < .001$) and Relationship Building (0.56, $p < .001$) dimensions. Likewise, the Relationship Maintenance and Relationship Building dimensions are positively correlated with each other (0.57, $P < .001$). Whereas, content consumption dimension is neutrally correlated with Content Transmission (0.09, $p = .003$) and Relationship Maintenance (.092, $p = .003$) and does not have any relationship with other interaction dimensions.

### 5.2.5 Distribution of Participants' Interactions with the Posts

This section presents the overall distribution of participants interactions with verified and unverified content in the control condition without addressing users' active-passive interactions tendencies. For each headline or post of this study, participants are given 4 options to select: Like, Comment, Share, and Take no action. We allow participants to select multiple options as our pilot study informs us that participants

like to use different interactions considering different scenarios. For instance, participant want to like or share the post if they see the post in Twitter, but would take no action for the same post if they see the post in Facebook. Participants mention that how they are connected with other persons in the social platform influences their interactions on that platform. In this study, we focus on users' general interaction tendencies across the social platforms they use.

In general, participants tend to use Like button when the content is verified. 503 participants see 4 true headlines and are allowed to choose 4 interaction options. Out of 2012 ($4 \times 503$) number of interactions, 849 interactions is for Like button, which is 42% of the all interactions for the verified post. After the Like button, participants use Share button for the verified post, which is 34% (688 out of 2012) of all the interactions. We find that the least used interaction of the verified post is the Comment button, 32% of the interactions.

In contrast, we find participants mostly take no action when they see the unverified posts - 45% (898 out of 2012) of the interaction decisions are Take no action for the unverified post. We find that participants want to use comment button more for the unverified post, which is 33% ( 662 out of 2012) of the interactions. After the comment button, participants tend to use Share button more than the Like button - 28% of the interactions are for share button and 26% of interactions are for Like button.

Table 5.6: Distribution of individuals' interactions with verified and unverified headlines.

| Headline ID | Like | Comment | Share | Take No Action |
|---|---|---|---|---|
| T1 | 141 (28%) | 162 (32%) | 167 (33%) | 210 (42%) |
| T2 | 182 (36%) | 166 (33%) | 161 (32%) | 188 (37%) |
| T3 | 272 (54%) | 158 (31%) | 174 (35%) | 127 (25%) |
| T4 | 254 (51%) | 162 (32%) | 186 (37%) | 134 (27%) |
| F1 | 114 (23%) | 160 (32%) | 137 (27%) | 240 (48%) |
| F2 | 129 (26%) | 172 (34%) | 144 (29%) | 216 (43%) |
| F3 | 127 (25%) | 179 (36%) | 138 (27%) | 227 (45%) |
| F4 | 152 (30%) | 151 (30%) | 142 (28%) | 215 (43%) |

### 5.2.6 Distribution of Participants' Perceptions of the Posts

This section presents participants' perceptions of the verified and unverified headlines selected in this study without addressing their active-passive interactions tendencies. Participants' in general find the selected verified headlines of this study contain helpful information - 73% of the responses report that the verified posts have helpful information. Similarly, participants find the selected unverified headlines contain harmful information - 62% of the responses report like that. Participants find the verified post more interesting and relevant to them in compare to the unverified post. 61% and 53% of responses report that the verified post seem interesting and relevant to the participants respectively, whereas 46% and 40% responses report that participants find the unverified post interesting and relevant to them respectively.

The highest 2 interactions for the headline T1 are Share and Take no action, and for the headline T2 are Like and Take no action. Participants find both of the headlines as helpful and interesting. When participants find headlines relevant, they tend to take actions. Participants find the headlines T3 and T4 relevant in addition to finding

the headlines as interesting and helpful. For these 2 headlines, participants selections of Take no action get reduced and the selections of using Like and Share button get increased.

Across the 4 unverified posts, the dominant interaction choices are Take no action and Comment button. On average, the Share interactions are higher than the Like interactions for these 4 posts. Though participants find the 4 unverified posts harmful, the responses of finding the headline F4 as helpful, interesting, and relevant are highest across the 4 unverified posts. We find that the selections of Like is also highest for F4 and closest to its Comment interaction. This indicates that how participants perceive the content influence their interactions with the unverified posts.

Table 5.7: Distribution of individuals' perceptions of verified and unverified headlines.

| **Headline ID** | **Helpful** | **Harmful** | **Interesting** | **Relevant** |
|---|---|---|---|---|
| T1 | 293 (58%) | 218 (43%) | 258 (51%) | 176 (35%) |
| T2 | 369 (73%) | 179 (36%) | 316 (63%) | 244 (49%) |
| T3 | 410 (82%) | 410 (82%) | 336 (67%) | 331 (66%) |
| T4 | 401 (79%) | 158 (31%) | 324 (64%) | 324 (64%) |
| F1 | 232 (46%) | 303 (60%) | 231 (46%) | 184 (37%) |
| F2 | 242 (48%) | 315 (63%) | 228 (45%) | 199 (40%) |
| F3 | 240 (48%) | 319 (63%) | 222 (44%) | 189 (38%) |
| F4 | 264 (52%) | 304 (60%) | 252 (50%) | 222 (44%) |

## 5.3 Data Analysis

This section discusses the process applied to investigate users' interaction tendencies in the active-passive continuum. The section shows the analysis used to find users' interactions with verified and unverified posts and the correlation between users' perception of content and their interaction with the content.

### 5.3.1 Clustering Users in the Active-Passive Continuum

To find the clusters of users in the active-passive continuum, we perform K-means clustering algorithms [64] on users' social media usage responses. We represent each response as a five-dimensional vector as participants provided their levels of agreement for five statements. We convert the options: Agree, Somewhat Agree, Neither Agree or Disagree, Somewhat Disagree, and Disagree into the numeric values 4, 3, 2, 1, 0, respectively - the higher numeric value indicates the higher agreement with the statement. After converting users' responses to the social media usage questionnaire into 5-dimensional numeric vectors, we train the k-mean algorithm on that vector representation.

We apply the elbow method to find the optimum number of clusters. We train the k-mean algorithm specifying the cluster number from 1 to 10 and calculate the distortions for all assigned clusters. Finally, we apply the Kneed algorithm [65] to identify the elbow point and the optimum number of clusters.

### 5.3.2 Analysis of Interactions Across the Clusters

The control condition has four options of interactions: Like, Comment, Share, and Take no action. The treatment condition includes the Like, Comment, and Take no action options. Additionally, the treatment condition has 4 and 3 sharing options for verified and unverified posts, respectively, instead of the one sharing option in the control condition. We classify these additional sharing options as one and compare the sharing differences between the control and treatment conditions. In addition, we compare participants' usage of 2 kinds of sharing: share-post and share-article, to investigate the difference between post and article sharing use in the treatment condition.

The clustering algorithm identifies 3 clusters of participants. We calculate the number of decisions taken for each interaction by the participants of 3 clusters. As

the participant numbers in the 3 clusters are different, we calculate the percentages of decisions taken for each interaction across the 3 participant clusters. We used Chi-square analysis to find how the participants across the 3 groups utilized the interaction functionalities and tested the statistical significance. For the decisions of each interaction option, we apply the chi-square to test whether the interaction decisions across 3 groups are independent or related to each other. As we perform the independence test for 3 categories, the degrees of freedom (df) are (3-1) = 2. We hypothesize that there exists a difference in the interactions across the 3 categories, and the null hypothesis is that there is no difference. When the p-value of the chi-square test is less than 0.05, we reject the null hypothesis and accept the alternative hypothesis.

### 5.3.3 Finding Correlations Between Content Perceptions and Interactions

The Pearson correlation analysis investigates the relationship between individuals' perception of the content and their usage of interaction functionalities. For each participant group, we calculated the number of helpful, harmful, interesting, or relevant was reported as 'yes' by participants and their number of decisions for each interaction. For instance, one participant saw four verified posts and four unverified posts and reported whether they found the posts helpful, harmful, interesting, or relevant - we calculated those numbers. We also calculated the number of interaction decisions made by that participant for those posts and created a matrix of perception and interaction for the participant. We perform a similar approach and create the matrix for all participants in the 3 clusters. Finally, we use that matrix to complete the Pearson correlation analysis between content perception and content interaction for 3 clusters of participants. The Pearson correlation analysis returns an r value indicating the linear relationship between -1 and 1; -1 indicates a positive correlation, +1 indicates a negative correlation, and 0 indicates no correlation. When the p-value was less than 0.05, we accepted the significance of the correlation.

Similarly, we use the Pearson correlation analysis to investigate the relationship between participants' level of trust in fact-checking journals and their decisions to interact with verified and unverified posts. The same correlation analysis is performed to find the relationship between participants' motivation to participate in combating misinformation and their interaction decisions.

In summary, this chapter describes a survey study containing a social media usage questionnaire to investigate users' active-passive interaction tendencies and study tasks to investigate users' interactions with verified and unverified posts. This chapter explains the designs and questionnaire pertain to the three design principles, how we select the headlines for the tasks, and participants' overall interactions and perceptions of those headlines. Finally, this chapter describes the analysis process performed in this study to get answers to the research questions.

# CHAPTER 6: THE EFFECT OF PERSONALIZED INTERVENTIONS ON USERS ACROSS THE ACTIVE PASSIVE CONTINUUM

This chapter presents the effect of personalized interventions on users across the active-passive continuum. This dissertation aims to show that users are different in terms of their usage of social media interaction functionalities, and personalized interaction-focused interventions can increase users' participation in combating misinformation. This research aims to answer the following four research questions:

RQ1: How do users across the active-passive continuum use the basic interaction functionalities (like, comment, share) differently when they are presented with the content's factual position (awareness design principle)?

RQ2: How do users' perceptions of content influence their usage of basic interaction functionalities across the active-passive continuum?

RQ3: How do users across the active-passive continuum participate in combating misinformation when they are provided with additional interaction functionalities (guidance design principle)?

RQ4: How do users across the active-passive continuum have preferences for the platform-based incentives that can increase their participation in combating misinformation (incentive design principle)?

This chapter begins describing how 3 clusters of participants have emerged on the active-passive continuum from participants' responses to the social media usage questionnaire. Then, the chapter presents the findings related to the first research

question: how do the participants in these 3 clusters use the basic interaction functionalities such as like, comment, and share for verified and unverified posts. Afterward, the chapter shows the influence of users' content perception on their usage of interaction functionalities, which are the findings related to the second research question. Followed by this chapter presents the findings related to the 3rd and fourth research questions and finally concludes the chapter with a summary of the results.

## 6.1 Three Clusters of Users on the Active-Passive Continuum

The 3 clusters of users have emerged from the analysis of the participant's (N=1006) responses to the social media usage questionnaire: active, moderately active, and passive. The centroids of 3 clusters [Table: 6.1] show that cluster number 1 has high values across five interaction dimensions, indicating that participants in cluster 1 mostly agreed with the five statements of the social media usage questionnaire. These responses are similar to the active users' social media usage tendencies, and we label cluster 1 as the active group. On the contrary, cluster number 3 has a high value in content consumption, but low values in the other four dimensions, indicating cluster 3 captures the participants with interaction tendencies of passive users. Thus, we label cluster 3 as the passive group. Finally, cluster number 2 has higher values than the passive group and lower values than the active group. Therefore, this cluster captures the participants who interacted with the content more than passive users but less than active users, and we label cluster 2 as the moderately active group. Similar 3 clusters of participants are identified in control (N=503) and treatment (N=503) conditions. However, the centroids capturing the participants responses in both conditions (N=1006) is used for this study analysis to categorize the participants of 2 conditions into the same 3 clusters.

Table 6.1: Three cluster centroids capture users' interaction tendencies as active, moderately active, and passive.

| | **Cluster 1** | **Cluster 2** | **Cluster 3** |
|---|---|---|---|
| Content Consumption (D1) | 4.42 | 4.12 | 4.47 |
| Content Creation (D2) | 4.40 | 2.81 | 1.36 |
| Content Transmission (D3) | 4.34 | 3.52 | 1.87 |
| Relationship Maintenance (D4) | 4.37 | 3.89 | 2.36 |
| Relationship Building (D5) | 4.36 | 3.07 | 1.56 |
| Group | Active | Moderately Active | Passive |

As expected, active users utilize the interaction functionalities more than the moderately active users, and the moderately active users utilize the interaction functionalities more than the passive users. Table 6.2 shows participants' interaction decisions for the eight posts (verified and unverified) presented in the control condition that adopts the awareness design principle. The interaction decisions for like, comment, and share functionalities decrease from active group to passive group. The percentage values of interaction decisions show the active group has the highest decisions for like, comment, and share. After the active group, the moderately active group has higher decisions for those three functionalities than the passive group. These interaction decision trends match participants' self-reported social media usage responses, showing that the participants in active, moderately active, and passive groups display their interaction decisions, respectively. A similar result is found for the participants in the treatment condition when the design adopts the guidance principle.

Table 6.2: Participants' interactions with posts decreases from active to passive in awareness principle.

| **Group and # of Participants** | **Like** | **Comment** | **Share** |
|---|---|---|---|
| Active (252 participants) | 927 (46%) | 975 (48%) | 945 (47%) |
| Moderately Active (161 participants) | 337 (26%) | 322 (25%) | 278 (22%) |
| Passive (90 participants) | 107 (15%) | 13 (2%) | 26 (4%) |

## 6.2 Interaction Differences Across the 3 Clusters in Awareness Principle

This section presents findings related to the first research question that investigates the differences across the 3 clusters regarding their interactions with verified and unverified posts. The findings analyze participants' interaction decisions of the control condition, where the design adopts the awareness principle and informs participants about the factual position of the posts.

Table 6.3: Awareness principle applied to unverified posts reduces passive and moderately active users' interactions more than active users.

| **Group** | **Posts** | **Like** | **Comment** | **Share** |
|---|---|---|---|---|
| Active | Verified | 528 (52%) | 489 (49%) | 489 (49%) |
| | Unverified | 399 (40%) | 486 (48%) | 456 (45%) |
| | Delta | -12%*** | -1% | -4% |
| Moderately Active | Verified | 233 (36%) | 154 (24%) | 178 (28%) |
| | Unverified | 104 (16%) | 168 (26%) | 100 (16%) |
| | Delta | -20%*** | +2% | -12%*** |
| Passive | Verified | 88 (24%) | 5 (1%) | 21 (6%) |
| | Unverified | 19 (5%) | 8 (2%) | 5 (1%) |
| | Delta | -19%*** | +1% | -5%*** |

*** $p < .001$; ** $p < .01$; * $p < .05$

The results show statistically significant differences across the 3 clusters regarding how participants decide to use the interaction functionalities for the verified posts ($p < 0.05$). Table 6.3 shows participants' usage of like, comment, and share functionalities for verified posts. The findings indicate that participants in the active group use the three basic interaction functionalities, like, comment, and share equally; 52%, 49%,

and 49%, respectively. In comparison, participants in the moderately active and passive groups prefer using the like functionality more than the comment or share functionalities. Similarly, both moderately active and passive groups utilize the share functionality of the verified posts more than the comment functionality.

In contrast, participants across three groups reduce their interactions with the unverified posts [Table 6.3]. However, the differences in how active, moderately active, and passive groups reduce their interactions have statistical significance; $p < 0.01$ in the chi-square test. The moderately active group reduces their usage of like and share functionalities more than the other two groups. After the moderately active group, participants of the passive group reduce their usage of like and share functionalities. Finally, the active group reduces their interactions least among the three groups, though users of this group remain active most on the social platforms.

## 6.3 Relationship between Content Perception and Interaction

This section presents findings related to the second research question that investigates how participants perceiving verified and unverified posts as helpful, harmful, interesting, or relevant influence their basic interaction functionality usage, such as like, comment, and share.

### 6.3.1 Perception of Verified Posts

Participants increase their interactions when they find the verified posts interesting, helpful, or relevant and decrease the interactions when they find the verified posts harmful. Table 6.4 shows the correlations between users' perception of the verified content and their usage of interactions with the verified content presented in the control condition that adopts the awareness principle. The table has the correlations of 4 perceptions with 9 variables; 3 participants groups $\times$ 3 basic interactions: like, comment, and share. Out of these 9 variables, there exist statistically significant positive correlations between interest and 8 variables. Afterward, 6 and 5 variables

positively correlate with helpful and relevant perceptions. Finally, the lowest number of positively correlated variables, 3, is found for the harmful perception.

Table 6.4: Relationship between perception and interaction with verified posts.

| **Perception of Verified Post** | **Groups** | **Like** | **Comment** | **Share** |
|---|---|---|---|---|
| Interesting | Active | $0.4^{***}$ | - | $0.2^{*}$ |
| | Moderately Active | $0.2^{*}$ | $0.2^{*}$ | $0.3^{***}$ |
| | Passive | $0.6^{***}$ | $0.3^{**}$ | $0.2^{*}$ |
| Relevant | Active | $0.3^{***}$ | - | - |
| | Moderately Active | $0.3^{***}$ | - | - |
| | Passive | $0.5^{***}$ | $0.2^{*}$ | $0.3^{**}$ |
| Helpful | Active | $0.2^{***}$ | $0.1^{*}$ | $0.2^{***}$ |
| | Moderately Active | $0.2^{**}$ | - | $0.2^{**}$ |
| | Passive | $0.5^{***}$ | - | - |
| Harmful | Active | $0.2^{**}$ | - | - |
| | Moderately Active | - | $0.5^{***}$ | $0.3^{***}$ |
| | Passive | - | - | - |

$^{***}$ $p < .001$; $^{**}$ $p < .01$; $^{*}$ $p < .05$

Active users use all the basic interaction functionalities: like, comment, and share functionalities when they find the verified posts helpful. Among the 4 perceptions, only the perception of finding the verified posts helpful influences active users' 3 interaction functionality usage (like: r=.2, p<.001; comment: r=.1, p<.05, share: r=.2, p<.001). In addition, active users use share functionality when they find the verified post interesting (r =.2, p<.05). Besides, active users use like functionality when they find verified posts interesting - active users have the strongest positive correlation in this relationship (r = .4, p <.0001).

Moderately active users have the strongest positive association between their usage

of comment functionality and finding a verified post harmful (r=.5, p<.0001). There is no other statistically significant correlation for comment functionality in the other 2 participants' groups. Additionally, the moderately active group participants utilize the share functionality when they find the verified posts harmful (r=.3, p<.001). This usage could mean that the moderately active participants utilize both comment and share functionalities to establish their opinion when they find the verified post harmful. Moreover, participants in that group utilize the share functionality when they find the verified post helpful (r=.2, p<.01) or interesting (r=.3, p <.001). Similarly, the moderately active group participants use the like functionality when they find the verified posts helpful (r=.2, p<.01), relevant (r=.3, p <.0001), or interesting (r=.2, p <.05).

Passive users have stronger positive correlations between the like button and when they find verified posts interesting (r=.6, p<.0001), relevant (r=.5, p<.0001), or helpful (r=.5, p<.0001). They also utilize the comment and share functionalities when they find the verified posts interesting or relevant. However, there are no other statistical correlations between passive users' finding verified posts helpful and their usage of comment or share functionalities.

### 6.3.2 Perception of Unverified Posts

Participants use comment functionality when they find unverified posts interesting, relevant, or helpful. Table 6.5 shows the correlations between participants' perception of unverified content and interaction usage. There are seven statistically significant positive correlations between comment functionality and unverified posts across the 3 clusters, whereas there exist 5 positive correlations between comment functionality and verified posts in Table 6.4. Besides, statistically significant negative correlations exist for the like and share functionalities when participants find the unverified posts harmful, indicating that participants reduce their usage of like and share buttons when they notice unverified posts contain harmful information. However, participants

across 3 groups increased their interactions with unverified posts when they found the unverified post interesting, helpful, or relevant.

Table 6.5: Relationship between perception and interaction with unverified posts.

| **Perception of Unverified Post** | **Groups** | **Like** | **Comment** | **Share** |
|---|---|---|---|---|
| Interesting | Active | 0.5*** | 0.2** | 0.3*** |
| | Moderately Active | 0.4*** | 0.3*** | 0.3*** |
| | Passive | 0.5*** | 0.3* | 0.4*** |
| Relevant | Active | 0.4*** | - | 0.3*** |
| | Moderately Active | 0.6*** | 0.3*** | 0.4*** |
| | Passive | 0.5*** | 0.3** | - |
| Helpful | Active | 0.5*** | 0.2** | 0.5*** |
| | Moderately Active | 0.6*** | 0.4*** | 0.5*** |
| | Passive | 0.3*** | - | 0.2** |
| Harmful | Active | - | - | -0.3*** |
| | Moderately Active | -0.3*** | - | -0.3** |
| | Passive | -0.3** | - | -0.3*** |

*** $p < .001$; ** $p < .01$; * $p < .05$

Active users have the strongest positive correlation for sharing unverified posts when they find them as helpful information (r=.5, p<.0001). Similarly, active users share unverified posts when they find the unverified information interesting (r=.3, p <.0001) or relevant (r=.3, p<.0001). However, active users decrease their sharing activities when they find unverified posts harmful (r=-.3, p <.0001). Again, active users have more positive correlations for comment functionality in unverified posts than in verified posts; 2 vs. 1 positive correlation. Similarly, active users use the like functionality when they find unverified posts helpful (r=.5, p<.0001), interesting (r=.5, p<.0001), or relevant (r=.4, p <.0001).

Moderately active users have similar statistical correlations for sharing functionality as active users. For example, moderately active users have the strongest positive correlation for share functionality when they find unverified posts helpful ($r=.5$, $p<.0001$). Similarly, a negative correlation exists when moderately active users find unverified posts harmful ($r=-.3$, $p<.01$). However, moderately active users have more positive correlations for comment functionality than active users; 3 vs. 2 correlations. In addition, moderately active users have negative correlations for like functionality when they find the unverified information harmful ($r=-.3$, $p<.001$) but exhibit positive correlations when they find unverified posts helpful ($r=.6$, $p<.0001$), interesting ($r=.4$, $p<.0001$), or relevant ($r=.6$, $p<.0001$).

Passive users share unverified posts when they find the unverified content helpful ($r=.2$, $p<.05$) or interesting ($r=.4$, $p<.001$). Similarly, passive users use like functionality for the unverified posts when they find the unverified information helpful ($r=.3$, $p<.001$), interesting ($r=.5$, $p<.0001$), or relevant ($r=.5$, $p<.0001$). However, passive users' usage of share and like functionalities negatively correlates when they find unverified posts harmful (share: $r=-.3$, $p<.001$; like: $r=-.3$, $p<.005$). Furthermore, passive users use comment functionality when they find unverified posts interesting ($r=.3$, $p<.05$) or relevant ($r=.3$, $p<.01$).

## 6.4 Interaction Differences between Awareness and Guidance Principles

This section shows the effectiveness of guidance principles in increasing users' participation in distributing credible information and compares the interaction differences between awareness and guidance principles.

### 6.4.1 Interaction Differences in Verified Posts between 2 Principles

Participants across the 3 groups increased their sharing functionality usage in the guidance design principle for the verified posts (the treatment condition); the difference in sharing usage between the two conditions is statistically significant across the

3 participants groups [Table 6.6]. For example, the active group participants use the share functionality 25% more in the treatment than in the control condition, with a statistical significance of p<.0001. Additionally, the moderately active group has a statistically significant increase in sharing usage (p<.0001), which is 20% more than the controlled condition. Similarly, passive users also display an increment in their usage of sharing functionalities, which is 13% more than the control group with a statistically significant of p<.0001. Notably, the statistically significant difference exists only for the usage of share functionality, and there are no statistically significant differences in the usage of other interaction functionalities, such as like and comment.

Table 6.6: Guidance principle facilitates the distribution of verified information more than the awareness principle.

| **Group** | **Condition (Verified Posts)** | **Like** | **Comment** | **Share** |
|---|---|---|---|---|
| Active | Awareness | 528 (52%) | 489 (49%) | 489 (49%) |
| | Guidance | 509 (51%) | 443 (44%) | 737 (74%) |
| | Delta | -1% | -5% | +25% *** |
| Moderately Active | Awareness | 233 (36%) | 154 (24%) | 178 (28%) |
| | Guidance | 249 (41%) | 131 (21%) | 291 (48%) |
| | Delta | +5% | -3% | +20% *** |
| Passive | Awareness | 88 (24%) | 5 (1%) | 21 (6%) |
| | Guidance | 118 (30%) | 13 (3%) | 75 (19%) |
| | Delta | +6% | +2% | +13% *** |

*** $p < .001$; ** $p < .01$; * $p < .05$

Active users increase their decisions to distribute credible information and utilize the verified post sharing and fact-checked article sharing functionalities presented in the treatment condition [Table 6.7]. For instance, active users in the treatment condition share the verified posts 14% more than the controlled condition (p<.0001).

In addition, active users utilize the fact-checked article sharing functionalities for 34% of the verified posts - the control condition does not include this article sharing functionality. The results also indicate that active users utilize the post sharing functionality 29% more than the article sharing functionality.

Table 6.7: Posts sharing and fact-checked article sharing of guidance principle applied to verified posts.

| **Group** | **Number of Interactions** | | **Remark** |
|---|---|---|---|
| Active | Share true post | 627 (63%) | - Guidance post sharing has increased than awareness post sharing by 14% *** |
| Active | Share fact-checked article | 342 (34%) | - Guidance post sharing has increase than guidance article sharing by 29% *** |
| Moderately Active | Share true post | 214 (35%) | - Guidance post sharing has increased than awareness post sharing by 7% ** |
| Moderately Active | Share fact-checked article | (165) 27% | - Guidance post sharing has increase than guidance article sharing by 8% ** |
| Passive | Share true post | (42) 11% | - Guidance post sharing has increased than awareness post sharing by 5% * |
| Passive | Share fact-checked article | (53) 13% | - Guidance article sharing has increased than guidance post sharing by 2% |

*** $p < .001$; ** $p < .01$; * $p < .05$

Moderately active users increase their decisions to distribute credible information

when they receive multiple functionalities in the treatment condition to share the verified posts [Table 6.7]. Compared to the controlled condition, moderately active users increased their decisions to share verified posts 7% more ($p<.01$). Additionally, the moderately active users utilize the fact-checked article sharing functionality presented in the treatment condition and decide to use the interaction functionality for 27% of the verified posts. Moreover, moderately active users utilize the post sharing functionality 8% more than the article sharing functionality.

Passive users utilize the fact-checked article sharing functionality most than the participants in active or moderately active groups. There is a 7% increased usage of this article sharing functionality than the traditional post sharing functionality of the controlled condition ($p<.0001$). Additionally, multiple post sharing functionalities facilitate a 5% additional sharing decisions of the passive users in the treatment condition compared to the controlled condition ($p<.05$).

### 6.4.2 Interaction Differences in Unverified Posts between 2 Principles

The interaction difference with unverified posts between awareness and guidance principle shows an increased usage of sharing functionality in the treatment condition that includes the fact-checked article sharing functionality [Table 6.8]. In the treatment condition, passive and moderately active participants utilize the fact-checked article sharing functionality more than the traditional post sharing functionality for the unverified posts [Table 6.9]. For example, passive participants share the fact-checked article 11% more than they share the unverified posts ($p<.0001$). Similarly, moderately active participants use the fact-checked article sharing functionality 7% more than the unverified post sharing functionality ($p<.01$). However, there is no significant difference in active participants' unverified post sharing and fact-checked article sharing usage.

Table 6.8: Interaction differences across 3 groups for unverified posts (awareness vs guidance).

| Group | Condition (Unverified Posts) | Like | Comment | Share |
|---|---|---|---|---|
| Active | Awareness | 399 (40%) | 486 (48%) | 456 (45%) |
| | Guidance | 412 (41%) | 459 (46%) | 650 (65%) |
| | Delta | +1% | -2% | +20% *** |
| Moderately Active | Awareness | 104 (16%) | 168 (26%) | 100 (16%) |
| | Guidance | 130 (21%) | 130 (21%) | 244 (40%) |
| | Delta | +5%* | -5% *** | +24% *** |
| Passive | Awareness | 19 (5%) | 8 (2%) | 5 (1%) |
| | Guidance | 13 (3%) | 20 (5.0%) | 70 (18%) |
| | Delta | -2% | +3% | +17% *** |

*** $p < .001$; ** $p < .01$; * $p < .05$

Passive users utilize the fact-checked article sharing functionality most - this additional functionality enables passive users to distribute the fact-checked article for 15% of the unverified posts. Usually, passive users interacted with the unverified posts less - they use the post sharing functionality in the controlled condition for 1% of the unverified posts. In comparison, the fact-checked article sharing functionalities increase passive users' participation for an additional 14% of the unverified posts and assist them in disturbing credible information, which has a statistical significance of $p < .0001$.

Table 6.9: Post and article sharing usage across 3 groups in guidance design principle.

| | **Active** | **Moderately Active** | **Passive** |
|---|---|---|---|
| Share verified posts | 627 (63%) | 214 (35%) | (42) 11% |
| Share unverified posts | 444 (44%) | 123 (20%) | 17 (4%) |
| Delta | -19%*** | -15%*** | -7%** |
| Share unverified post | 444 (44%) | 123 (20%) | 17 (4%) |
| Share fact-checked article when posts are unverified | 425 (43%) | 165 (27%) | 60 (15%) |
| Delta | -1% | +7%** | +11%*** |
| Share fact-checked article when posts are verified | 342 (34%) | (165) 27% | (53) 13% |
| Share fact-checked article when posts are unverified | 425 (43%) | 165 (27%) | 60 (15%) |
| Delta | +9%** | 0% | +2% |

*** $p < .001$; ** $p < .01$; * $p < .05$

Moderately active users utilize the fact-checked article sharing functionality for 27% of the unverified posts. This distribution of credible information is 11% more than the usage of unverified post sharing functionality in the controlled condition ($p<.0001$). In addition, moderately active users distribute the fact-checked article 7% more than their sharing of unverified posts in the treatment condition ($p < .01$).

Active users utilize the fact-checked article sharing functionality 9% more when posts are unverified than verified posts ($p<.001$). In contrast, moderately active and passive groups similarly use the fact-checked article sharing functionalities for unverified and verified posts; there is no significant difference in how the two groups utilize the article sharing functionalities for verified and unverified posts.

The differences between post sharing and fact-checked article sharing usage when posts are verified and unverified in the guidance principle are presented in table 6.10. In the guidance design principle (treatment condition), active, moderately active, and passive users share unverified posts 19%, 15%, and 7% less in comparison to their sharing decisions of the verified posts ($p<.01$). However, moderately active and passive groups exhibit a 4% and 3% increase respectively in sharing unverified posts in the treatment condition than in the controlled condition. As participants in the treatment often use fact-checked article sharing and unverified post sharing functionalities for identical posts - this sharing usage could indicate that some participants used both information to justify their points.

## 6.5 Platform-based Incentives and Motivation Levels to Make the Truth Louder

This section presents the findings of the fourth research question that investigates the platform-based incentive preferences that can appeal to users with various interaction tendencies. The section shows that participants' motivation levels for combating misinformation and their trust in fact-checking journals influence users' interaction decisions with verified and unverified posts.

### 6.5.1 Preference for Platform-based Incentives

Participants' across the active-passive continuum show different preferences for the platform-base incentives that can inspire their participation for combating misinformation [Table 6.10]. Active participants in both conditions show higher levels of preference for getting badges (57%) and followers (53%). In contrast, participants in the passive group report lower levels of preference for those 2 incentives, 22% and 12% respectively. The moderately active participants, similar to the participants in the passive group, have lower percentage of responses for the 2 incentives, 34% and 29% respectively.

Participants in the passive group report higher levels of preference for getting in-

formation regarding the impact they are making; how their participation is helping other social media users. Among the four platform-based incentives, participants in the passive group have the highest percentage of responses for the incentive, which is 47%. Conversely, active participants display the lowest preference for the incentive, which is 33%. However, the moderately active participants exhibit similar preferences to the passive group - 43% of their responses are for this incentive. In addition, the moderately active participants have a higher preference for another incentive, content prioritization, which prioritizes individuals' content to other users (47%). Besides, moderately active users are inclined to get badges and followers as incentives. These similar trends exist in both control and treatment conditions.

Table 6.10: Participants across the active-passive continuum exhibit different preference toward platform-based incentives.

| **Platform-based incentives** | **Statements of the incentive** | **Active** | **Moderately Active** | **Passive** |
|---|---|---|---|---|
| Getting badges | "The platform gives me badges that inform other users about my contribution for combating misinformation." | 287 (57%) | 106 (34%) | 42 (22%) |
| Getting followers | "The platform suggests other users to follow my account as I contribute in combating misinformation." | 268 (53%) | 92 (29%) | 23 (12%) |
| Content prioritization | "The platform prioritizes my posts to other users as I contribute in combating misinformation." | 234 (47%) | 126 (40%) | 45 (24%) |
| Receive information regarding the impact | "The platform shows me how I am helping my friends and community by participating in combating misinformation." | 165 (33%) | 134 (43%) | 90 (47%) |
| Other | "Other" | 10 (2%) | 36 (12%) | 55 (29%) |

### 6.5.2 Participants' Motivation Levels vs. Participants' Interactions with Posts

Passive users who report their motivation of spreading credible information use like and share functionalities for the verified posts (like: $r=.4$, $p<.001$; share: $r=.3$; $p<.01$). Similarly, when the posts are unverified in the guidance design principle, there exists a positive relationship between passive users' motivation and their usage

of fact-checked article sharing (r= 0.2; p<.05). These findings show the significance of designing interactions for passive users that facilitate motivate passive users to participate in spreading credible information.

The moderately active participants motivated to spread credible information use the like functionally for verified posts (r=.2; p<.05). Moreover, the moderately active participants motivated to mitigate the spread of misinformation reduce their usage of like and sharing functionalities for the unverified posts (like: r=-.2; p<.05; share: r=-.2; p<.01). However, there is no statistically significant correlation between active users' motivation levels and their usage of interactions.These findings suggest that active users, due to their interaction tendencies, have less control over their interactions with the content than passive and moderately active participants.

### 6.5.3 Trust in Fact-checking Journals vs. Interactions with Posts

Participants' level of trust for the credibility judgment made by the fact-checking journals influences their interactions with verified and unverified posts [Table 6.11]. Active users who trust fact-checking journals use like, comment, and share functionalities for the verified posts (like:r=.3, p<.0001; comment: r=.2, p< .05; share: r=.2, p<.01). However, active users who trust fact-checking journals do not reduce their like and comment uses for unverified posts (like: r=.2; p<.001; comment: r=.2; p<.001). In addition, there is no other statistically significant correlation between active users' share functionality usage for the unverified posts and their level of trust in the fact-checking journals.

Table 6.11: Correlation between participants' trust in fact-checking journals and their interactions with posts.

| Post | Groups | Like | Comment | Share |
|---|---|---|---|---|
| Verified | Active | $0.3^{***}$ | $0.2^{*}$ | $0.2^{**}$ |
| | Moderately Active | $0.3^{***}$ | - | - |
| | Passive | $0.3^{**}$ | - | - |
| Unverified | Active | $0.2^{***}$ | $0.2^{***}$ | - |
| | Moderately Active | - | - | - |
| | Passive | - | - | $-0.2^{*}$ |

$^{***}$ $p < .001$; $^{**}$ $p < .01$; $^{*}$ $p < .05$

Conversely, passive participants who trust fact-checking journals reduce their sharing when posts are unverified (r=-.2; p<.05). Moreover, passive and moderately active participants like the verified posts when they trust fact-checking journals (passive: r=.3, p<.01; moderately active: r=.3, p<.0001). These findings indicate that users across the active-passive continuum utilize the interaction functionalities differently when they trust fact-checking journals.

In the guidance design principle, when passive and moderately active participants receive additional sharing functionalities, participants who trust the fact-checking journals increase their sharing of verified information. For instance, passive users and moderately active users have statistically positive correlation between their trust in fact-checking journals and their usage of sharing functionalities in the guidance principle (passive: r =.3, p<.001; moderately active: r =.3, p<.001). There are no significant differences for the active participants. These findings suggest the effectiveness of the guidance principle for facilitating passive and moderately active users' participation in distributing credible information.

## 6.6 Summary

This research focuses on determining interaction patterns across the users due to their interaction tendencies on social media - how users with different interaction tendencies interact with verified and unverified posts when they become aware of the factual position of the content. This study investigates the users across the active-passive continuum regarding their usage of basic interaction functionalities such as like, comment, and share; how their perception of the content influences the use of interaction functionalities. In addition, the study examines the effect of the additional supportive interaction functionalities on those users and their preference for platform-based incentives that intend to increase user participation in combating misinformation.

The results indicate that the usage of like, comment, and share functionalities decrease from active to passive for the verified posts. Participants across the active-passive continuum prefer to use like functionalities most for the verified posts; after that, they utilize the share functionality, and finally, they utilize the comment functionality least. Participants across the continuum tend to interact less with the unverified posts. However, participants who display passive user tendencies reduce their interaction with unverified posts more than those who display active user tendencies.

The study reveals that participants across the active-passive continuum interact with unverified posts when they find the unverified information helpful, relevant, or interesting. At the same time, participants reduce their interactions with the unverified posts when they perceive the content as harmful information. Moreover, the results show a similar relationship between users' perception and their interaction with the verified posts - participants interact with the verified posts when they find the information helpful, relevant, or interesting, and reduce their interactions with the verified post when they identify the verified posts as harmful. Though participants across the active-passive continuum display similar trends for verified and unverified

posts, their usage of the interaction functionalities such as like, comment, and share differ based on the interaction tendencies participants possess on the active-passive continuum.

The results show that participants increase their participation in distributing credible information when they receive supportive interaction functionalities. Furthermore, participants across the active-passive continuum increase their usage of sharing functionalities in the guidance design principle than in the awareness design principle. However, the guidance design principle's effect differs across the active-passive continuum. For instance, the results reveal that participants with passive user tendencies utilize the sharing functionality of the fact-checked article for both verified and unverified posts. Conversely, participants with active user tendencies utilize the traditional post sharing functionality for the verified posts and the fact-checked article sharing functionality for the unverified posts.

Participants across the active-passive continuum reveal the opposite preferences toward platforms-based incentives that can inspire their participation in combating misinformation. For example, participants displaying active user tendencies show preferences for extrinsic rewards such as getting badges or having followers on the social platforms that can inform other social media users about active users' contributions to combating misinformation. Conversely, participants showing passive user tendencies indicate a preference for the intrinsic rewards, such as getting educated about the impact their participation can make on their friends and community.

This chapter presents the results related to the four research questions that show participants use the interaction functionality differently due to their interaction tendencies and how they perceive the content influence their interaction choices. Moreover, the three design principles affect the users with different interaction tendencies, increase their participation in distributing credible information, and reveal their differences in preference for platform-based incentives. These interaction-focused find-

ings can contribute to designing platform-based interventions personalized to users' interaction tendencies.

# CHAPTER 7: DISCUSSION AND CONCLUSION

This dissertation investigates the potential of a new research direction that can mitigate misinformation by making the truth louder. This chapter discusses the key findings, suggests recommendations for personalized interaction-focused interventions, and direction for future research. Finally, this chapter finishes with a conclusion of the dissertation.

## 7.1 Interpretation and Discussion

This dissertation shows that users use the interaction functionalities differently due to their interaction tendencies on social platforms. In addition to users' post-sharing usage, this research investigates users' other interaction usage, such as like and comment functionalities, which previous studies [6, 9, 7] have not explored. This study reveals the difference in users' interaction preferences that designers can address to develop personalized interaction-focused intervention.

This research indicates that people prioritize their perception of finding posts helpful or harmful over the posts' factual positions determined by fact-checking journals. For example, people tend to interact more with unverified posts when they see them as helpful information. Similarly, people tend to interact less with verified posts when they think the post contains harmful information. Though previous research [6] show that users share interesting and relevant posts, these studies do not investigate how the perceptions of finding post helpful or harmful can influence the interactions. This research contributes to the existing literature by highlighting the importance of the perception: finding posts as helpful or harmful, to influence users' interactions with verified and unverified content. These findings can contribute to designing personal-

ized interventions and combat misinformation.

In this study, three kinds of social media usage have emerged from participants' responses to the social media usage questionnaire. The first type of usage prefers creating and distributing content. The second type of usage utilizes social media's relationship-building and maintenance elements in addition to creating and distributing content. Finally, the third type of usage solely focuses on content consumption. These three types of social media usage align with the three kinds of engagement on Facebook studied in [13]. Gerson et al. [13] have labeled these three types of engagement as active non-social, active social, and passive, where active non-social refers to engagement necessary to produce and share content without displaying any social nature on the platform, such as interaction with other users. Conversely, active social engagement refers to the platform's active users that display a social nature, such as commenting and chatting with friends. Finally, passive engagement refers to the content consummation nature of the social media users. However, this dissertation distinguishes users based on their social media interaction tendencies and identifies 3 clusters of interaction tendencies across the active-passive continuum: active, moderately active, and passive.

The three clusters of interaction tendencies: active, moderately active, and passive, is the basis of personalized intervention in this study. The active users possess the highest interaction tendencies, the moderately active users possess the medium interaction tendencies, and the passive users display the lowest interaction tendencies of the active-passive continuum. This dissertation shows the effectiveness of the guidance principle for assisting users in distributing credible information when users possess lower interaction tendencies as passive and moderately active users. For example, moderately active and passive users increase their usage of sharing functionalities in the design that adopts the guidance principle compared to the design that adopts the awareness principle. In contrast, the active users do not have any increased usage

of share functionality in the guidance compared to the awareness principle. Hence, the guidance principle is the personalized intervention for the moderately active and passive users when the design goal is to increase their participation in distributing credible content.

The previous literature supports that passive and moderately active users can increase their participation when they receive interaction support. Shao [37] suggests that users who consume content and have lower participation in social media possess the talent to increase their participation and become content creators on the platform. Preece et al. [17] highlight the importance of design support that transforms a person from reader to leader in online activities. Fogg [59] advocates that individuals with lower ability can increase their ability to perform tasks when they receive the necessary support. This dissertation shows that the design principle addressing users' interaction ability can assist users in distributing credible information on social platforms.

On the other hand, active users possess the highest interaction tendencies and participate in distributing credible content. However, due to active users' highest interaction tendencies, active users also interact more with unverified posts than moderately active and passive users. Gerson et al. [13] have found that Facebook users who engaged in an active non-social way display an impulsivity trait, which refers to individuals' inclination to disinhibited and unplanned behavior. This impulsivity trait could be a reason that explains why active users can not limit their interactions with unverified posts. Thus, the personalized intervention for active users should be different and focus on assisting active users in reducing their interactions with unverified posts.

Additionally, this dissertation reveals users' various preferences for the platform-based incentives that can encourage them to participate in combating misinformation. For example, active users are interested in getting badges and have increased followers.

These platform-based incentives attract active users as these advantages increase their visibility and credibility on social platforms. However, passive users display lower attraction for these incentives as they do not prefer having any digital footprint. This study shows that passive users who report their motivation to spread credible information utilize the sharing functionalities to distribute verified information. In addition, passive users prefer to make an impact on their friends and community and want to learn more about the ways they can make the impact. The moderately active users' preferences for the incentives are primarily similar to passive users, but they also display incentive preferences as the active users. Thus, personalized intervention should address the various preference in the active-passive continuum when designing platform-based incentives to encourage users to combat misinformation.

## 7.2 Limitations

This study has limitations because of conducting a survey study to collect participants' self-reported interaction decisions to the social media posts. The results of this study are yet to be generalized, which do not study users' responses to more than one topic and do not address hidden factors or biases that might influence the interaction decisions.

This research is limited in the generalizability of the results due to focusing on only one topic: the Covid-19 vaccine. The study chooses one topic to eliminate the influence of confounding variables while studying participants' interactions with verified and unverified posts. Future research can investigate the effect of design principles on additional topics and examine the generalizability of the result.

Though the headlines of this study intend to reduce political biases, due to the politically-charged nature of the Covid-19 topic, individuals might have preceding beliefs and motivations that could drive interactions, which this study does not address. Future research can study the correlation between individuals' misinformation-sharing motivations in political settings and participants' responses to the design principles

applied to politically-biased posts.

This research focuses on users' general social media usage and interaction tendencies, not the usage or interactions particular to any social media platform. This experiment includes tasks as a survey study that does not address the affordance or biases on existing social media platforms. This study collects participants from Amazon Mechanical Turk, not any social media platform, though the participants of this study are users of various social media platforms.

In summary, the results of this study lack generalizability due to selecting only one topic to study participants' interaction behavior. The result does not address the affordance or biases of the existing social media platforms as the results are collected from the survey study.

## 7.3 Future Work

This dissertation explores a novel research direction to mitigate the spread of misinformation by leveraging users' interaction tendencies so that users increase their interactions with verified posts and decrease their interactions with unverified posts. This dissertation prepares the foundation for future research to investigate the difference between social media users due to their interaction tendencies and use the findings to develop interventions personalized to users' interactions. Additional research is needed to correlate users' motivations to share misinformation with their interaction tendencies and study the effect of design principles when people have political biases. Furthermore, additional research can develop browser extensions and study users' actual behaviors on social platforms instead of collecting participants' self-reported responses to the survey studies.

Future research should study the correlation between individuals' motivation to share misinformation and their responses to the design principles that address users' active-passive tendencies. When participants are exposed to the headlines, this study asks whether participants find the posts interesting, relevant, helpful, or harmful

and correlates the results with their active-passive interaction tendencies. There are many other reasons why people share misinformation found in [cite]. Future studies can add additional motivations as options for the participants to choose and collect information about why individuals interact with the headlines. This information can correlate individuals' motivations to share misinformation and their active-passive interaction tendencies.

Future research should study the generalizability of this study's results by conducting survey studies on different topics, including politically biased topics. Future research can select headlines on topics other than the Covid-19 vaccine and use the same survey study to investigate participants' responses to the design principles. Future research can select political and controversial headlines and use the survey study used in this research to investigate the effect of design principles when people possess certain political views. These future studies can collect information regarding participants' political positions and correlate participants' political views with their active-passive interaction tendencies.

Additional in-depth research is needed to understand what platform-based incentives can motivate people to participate in making the truth louder and mitigating the spread of misinformation. This study suggests that users are willing to participate in making the truth louder and have preferences for platform-based incentives that encourage their participation on social platforms. Future research should conduct interviews with active and passive users and investigate the themes that emerged from their responses. Future research findings can reveal additional information about participants' motivations and requirements from the platforms to increase their participation in making the truth louder.

Future research can develop browser extensions to study users' actual social media behavior on social platforms rather than a survey of self-reported behavior. Future studies with the browser extension can present users with the UI adopting the 3 design

principles and investigate the participant's responses to the design principles. Studies can analyze social media data that includes users' interaction-related data, categorize users based on their interaction tendencies, find the type of topics individuals prefer to interact with, and correlate the findings with users' responses to the design principles. Future research should develop alternative design instances following the same design principles so that browsers can present effective UI designs in response to users' active-passive interaction tendencies.

In summary, future research can extend this research by studying the correlation between individuals' motivation and political biases with their active-passive tendencies and investigating the generability of this study's results. Future research can develop browsers extension instead of survey studies to investigate the effectiveness of the design principles on the users engaged in social media platforms.

## 7.4 Conclusion

This dissertation develops a foundation for the personalized interaction-focused intervention that leverages users' interaction tendencies to make the truth louder and mitigate the spread of misinformation. This study identifies three clusters of social media users based on their interaction tendencies: active, moderately active, and passive, where active users possess higher interaction tendencies than moderately active users, and moderately active users possess higher interaction tendencies than passive users. This dissertation addresses the differences between the interaction tendencies across three clusters and develops three principles of social media interactions that can assist users in combating misinformation.

A survey study with 1006 participants indicates that moderately active and passive users increase their participation when they receive additional interaction support and utilize the interaction functionalities to distribute credible information. Moreover, active, moderately active, and passive users show various preferences for platform-based incentives that can motivate them to participate more in combating misinformation.

Active users prefer platform-based incentives, such as getting badges or having the advantage on the platform that gets followers, whereas passive users want information from platforms regarding the impact of their participation on their friends and community. Moderately active users favor platform-based incentives as passive users and exhibit preference as active users. Additional research is needed to develop effective personalized interaction-focused interventions that can transform users' long-term interaction behaviors so that social media users increase their interaction with verified information and reduce their interactions with unverified information.

# REFERENCES


[1] S. Siddiqui and M. L. Maher, "Reframing the fake news problem: Social media interaction design to make the truth louder.," in *CHIRA*, pp. 158–165, 2021.

[2] S. Siddiqui and M. L. Maher, "Active-passive framework for developing communication strategies to combat misinformation.," in *Proceedings of the 19th International Conference on Web Based Communities and Social Media*, 2022.

[3] S. Lewandowsky, U. K. Ecker, and J. Cook, "Beyond misinformation: Understanding and coping with the "post-truth" era," *Journal of Applied Research in Memory and Cognition*, vol. 6, pp. 353–369, 2017.

[4] J. Smith, "Designing against misinformation." https://medium.com/facebook-design/designing-against-misinformation-e5846b3aa1e2, Last retrieved January 7, 2021.

[5] Y. Roth and N. Pickles, "Updating our approach to misleading information." https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html, Last retrieved January 7, 2021.

[6] W. Yaqub, O. Kakhidze, M. L. Brockman, N. Memon, and S. Patil, "Effects of credibility indicators on social media news sharing intent," in *Proceedings of the 2020 chi conference on human factors in computing systems*, pp. 1–14, 2020.

[7] E. Nekmat, "Nudge effect of fact-check alerts: Source influence and media skepticism on sharing of news misinformation in social media," *Social Media+ Society*, vol. 6, no. 1, p. 2056305119897322, 2020.

[8] M. M. Bhuiyan, K. Zhang, K. Vick, M. A. Horning, and T. Mitra, "Feedreflect: A tool for nudging users to assess news credibility on twitter," in *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 205–208, 2018.

[9] G. Pennycook, J. McPhetres, Y. Zhang, J. G. Lu, and D. G. Rand, "Fighting covid-19 misinformation on social media: experimental evidence for a scalable accuracy-nudge intervention," *Psychological science*, vol. 31, no. 7, pp. 770–780, 2020.

[10] M. Flintham, C. Karner, K. Bachour, H. Creswick, N. Gupta, and S. Moran, "Falling for fake news: investigating the consumption of news via social media," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 376, ACM, 2018.

[11] A. Chen, Y. Lu, P. Y. Chau, and S. Gupta, "Classifying, measuring, and predicting users' overall active behavior on social networking sites," *Journal of Management Information Systems*, vol. 31, no. 3, pp. 213–253, 2014.

[12] B. M. Trifiro and J. Gerson, "Social media usage patterns: Research note regarding the lack of universal validated measures for active and passive use," *Social Media+ Society*, vol. 5, no. 2, p. 2056305119848743, 2019.

[13] J. Gerson, A. C. Plagnol, and P. J. Corr, "Passive and active facebook use measure (paum): Validation and relationship to the reinforcement sensitivity theory," *Personality and Individual Differences*, vol. 117, pp. 81–90, 2017.

[14] C. Geeng, S. Yee, and F. Roesner, "Fake news on facebook and twitter: Investigating how people (don't) investigate," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2020.

[15] S. Lewandowsky, U. K. Ecker, C. M. Seifert, N. Schwarz, and J. Cook, "Misinformation and its correction: Continued influence and successful debiasing," *Psychological science in the public interest*, vol. 13, no. 3, pp. 106–131, 2012.

[16] N. Mele, D. Lazer, M. Baum, N. Grinberg, L. Friedland, K. Joseph, W. Hobbs, and C. Mattsson, "Combating fake news: An agenda for research and action," *Retrieved on October*, vol. 17, p. 2018, 2017.

[17] J. Preece and B. Shneiderman, "The reader-to-leader framework: Motivating technology-mediated social participation," *AIS transactions on human-computer interaction*, vol. 1, no. 1, pp. 13–32, 2009.

[18] S. Kumar and N. Shah, "False information on web and social media: A survey," *arXiv preprint arXiv:1804.08559*, 2018.

[19] B. D. Horne and S. Adali, "This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news," in *Eleventh International AAAI Conference on Web and Social Media*, 2017.

[20] S. Kumar, R. West, and J. Leskovec, "Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes," in *Proceedings of the 25th international conference on World Wide Web*, pp. 591–602, International World Wide Web Conferences Steering Committee, 2016.

[21] V. Rubin, N. Conroy, Y. Chen, and S. Cornwell, "Fake news or truth? using satirical cues to detect potentially misleading news," in *Proceedings of the second workshop on computational approaches to deception detection*, pp. 7–17, 2016.

[22] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," *arXiv preprint arXiv:1708.07104*, 2017.

[23] W. Y. Wang, ""liar, liar pants on fire": A new benchmark dataset for fake news detection," *arXiv preprint arXiv:1705.00648*, 2017.

[24] A. Arif, K. Shanahan, F.-J. Chou, Y. Dosouto, K. Starbird, and E. S. Spiro, "How information snowballs: Exploring the role of exposure in online rumor propagation," in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pp. 466–477, ACM, 2016.

[25] K. Starbird, "Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on twitter," in *Eleventh International AAAI Conference on Web and Social Media*, 2017.

[26] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei, "Rumor has it: Identifying misinformation in microblogs," in *Proceedings of the conference on empirical methods in natural language processing*, pp. 1589–1599, Association for Computational Linguistics, 2011.

[27] F. Jin, E. Dougherty, P. Saraf, Y. Cao, and N. Ramakrishnan, "Epidemiological modeling of news and rumors on twitter," in *Proceedings of the 7th Workshop on Social Network Mining and Analysis*, p. 8, ACM, 2013.

[28] S. L. Van der Linden, A. A. Leiserowitz, G. D. Feinberg, and E. W. Maibach, "How to communicate the scientific consensus on climate change: plain facts, pie charts or metaphors?," *Climatic Change*, vol. 126, no. 1-2, pp. 255–262, 2014.

[29] S. Van der Linden, A. Leiserowitz, S. Rosenthal, and E. Maibach, "Inoculating the public against misinformation about climate change," *Global Challenges*, vol. 1, no. 2, p. 1600008, 2017.

[30] J. Cook, S. Lewandowsky, and U. K. Ecker, "Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence," *PloS one*, vol. 12, no. 5, p. e0175799, 2017.

[31] R. H. Thaler and C. R. Sunstein, *Nudge: Improving decisions about health, wealth, and happiness*. Penguin, 2009.

[32] A. Acquisti, I. Adjerid, R. Balebako, L. Brandimarte, L. F. Cranor, S. Komanduri, P. G. Leon, N. Sadeh, F. Schaub, M. Sleeper, *et al.*, "Nudges for privacy and security: Understanding and assisting usersâ choices online," *ACM Computing Surveys (CSUR)*, vol. 50, no. 3, pp. 1–41, 2017.

[33] G. Pennycook, Z. Epstein, M. Mosleh, A. A. Arechar, D. Eckles, and D. Rand, "Understanding and reducing the spread of misinformation online," *Unpublished manuscript: https://psyarxiv. com/3n9u8*, 2019.

[34] M. Mosleh, G. Pennycook, and D. G. Rand, "Self-reported willingness to share political news articles in online surveys correlates with actual sharing on twitter," *Plos one*, vol. 15, no. 2, p. e0228882, 2020.

[35] V. o. I. Guy Rosen and H. o. N. F. I. Tessa Lyons, "Remove, reduce, inform: New steps to manage problematic content." https://about.fb.com/news/2019/04/remove-reduce-inform-new-steps/, Last retrieved January 7, 2021.

[36] S. Fiegerman, "Facebook, google, twitter to fight fake news with 'trust indicators'." https://money.cnn.com/2017/11/16/technology/tech-trust-indicators/index.htm, Last retrieved January 7, 2021.

[37] G. Shao, "Understanding the appeal of user-generated media: a uses and gratification perspective," *Internet research*, vol. 19, no. 1, pp. 7–25, 2009.

[38] C. Wardle, "Fake news. it's complicated," *First Draft News*, vol. 16, 2017.

[39] E. C. Tandoc Jr, Z. W. Lim, and R. Ling, "Defining "fake news" a typology of scholarly definitions," *Digital journalism*, vol. 6, no. 2, pp. 137–153, 2018.

[40] G.-W. Bock, R. W. Zmud, Y.-G. Kim, J.-N. Lee, *et al.*, "Behavioral intention formation in knowledge sharing: Examining the roles of extrinsic motivators, social-psychological factors, and organizational climate.," *MIS quarterly*, vol. 29, no. 1, pp. 87–111, 2005.

[41] P. Blau, *Exchange and power in social life*. Routledge, 2017.

[42] R. J. Vallerand, "Toward a hierarchical model of intrinsic and extrinsic motivation," in *Advances in experimental social psychology*, vol. 29, pp. 271–360, Elsevier, 1997.

[43] A. E. Marwick, "Why do people share fake news? a sociotechnical model of media effects," *Georgetown Law Technology Review*, vol. 2, no. 2, pp. 474–512, 2018.

[44] B. Osatuyi, "Information sharing on social media sites," *Computers in Human Behavior*, vol. 29, no. 6, pp. 2622–2631, 2013.

[45] S. Chaiken, "Heuristic and systematic information processing within and beyond the persuasion context," *Unintended thought*, pp. 212–252, 1989.

[46] L. Liu, J. Tang, J. Han, and S. Yang, "Learning influence from heterogeneous social networks," *Data Mining and Knowledge Discovery*, vol. 25, no. 3, pp. 511–544, 2012.

[47] X. Chen, S.-C. J. Sin, Y.-L. Theng, and C. S. Lee, "Why students share misinformation on social media: Motivation, gender, and study-level differences," *The Journal of Academic Librarianship*, vol. 41, no. 5, pp. 583–592, 2015.

[48] F. Bentley, K. Quehl, J. Wirfs-Brock, and M. Bica, "Understanding online news behaviors," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–11, 2019.

[49] R. Torres, N. Gerhart, and A. Negahban, "Combating fake news: An investigation of information verification behaviors on social networking sites," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.

[50] B. Swire, A. J. Berinsky, S. Lewandowsky, and U. K. Ecker, "Processing political misinformation: comprehending the trump phenomenon," *Royal Society open science*, vol. 4, no. 3, p. 160802, 2017.

[51] M. J. Metzger, A. J. Flanagin, and R. B. Medders, "Social and heuristic approaches to credibility evaluation online," *Journal of communication*, vol. 60, no. 3, pp. 413–439, 2010.

[52] A. J. Berinsky, "Rumors and health care reform: Experiments in political misinformation," *British journal of political science*, vol. 47, no. 2, pp. 241–262, 2017.

[53] C. R. Sunstein, S. Bobadilla-Suarez, S. C. Lazzaro, and T. Sharot, "How people update beliefs about climate change: Good news and bad news," *Cornell L. Rev.*, vol. 102, p. 1431, 2016.

[54] P. M. Sara Su, "New test with related articles." https://about.fb.com/news/2017/04/news-feed-fyi-new-test-with-related-articles/, Last retrieved January 7, 2021.

[55] N. F. Adam Mosseri, VP, "Addressing hoaxes and fake news." https://about.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/, Last retrieved January 7, 2021.

[56] S. Siddiqui, M. L. Maher, N. Najjar, M. Mohseni, and K. Grace, "Personalized curiosity engine (pique): A curiosity inspiring cognitive system for student directed learning.," in *CSEDU (1)*, pp. 17–28, 2022.

[57] K. M. Greenhill and B. Oppenheim, "Rumor has it: The adoption of unverified information in conflict zones," *International Studies Quarterly*, vol. 61, no. 3, pp. 660–676, 2017.

[58] L. Fazio, "Pausing to consider why a headline is true or false can help reduce the sharing of false news," *Harvard Kennedy School Misinformation Review*, vol. 1, no. 2, 2020.

[59] B. J. Fogg, "A behavior model for persuasive design," in *Proceedings of the 4th international Conference on Persuasive Technology*, pp. 1–7, 2009.

[60] "New test with related articles." https://www.politifact.com/, Last retrieved January 7, 2021.

[61] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *psychometrika*, vol. 16, no. 3, pp. 297–334, 1951.

[62] G. Ursachi, I. A. Horodnic, and A. Zait, "How reliable are measurement scales? external factors with indirect influence on reliability estimators," *Procedia Economics and Finance*, vol. 20, pp. 679–686, 2015.

[63] C. Hulin, R. Netemeyer, and R. Cudeck, "Can a reliability coefficient be too high?," *Journal of Consumer Psychology*, vol. 10, no. 1/2, pp. 55–58, 2001.

[64] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," *Pattern recognition*, vol. 36, no. 2, pp. 451–461, 2003.

[65] V. Satopaa, J. Albrecht, D. Irwin, and B. Raghavan, "Finding a" kneedle" in a haystack: Detecting knee points in system behavior," in *2011 31st international conference on distributed computing systems workshops*, pp. 166–171, IEEE, 2011.