International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hihc20

“It is Luring You to Click on the Link With False Advertising” - Mental Models of Clickbait and Its Impact on User’s Perceptions and Behavior Towards Clickbait Warnings

Ankit Shrestha, Arezou Behfar & Mahdi Nasrullah Al-Ameen

To cite this article: Ankit Shrestha, Arezou Behfar & Mahdi Nasrullah Al-Ameen (08 Mar 2024): “It is Luring You to Click on the Link With False Advertising” - Mental Models of Clickbait and Its Impact on User’s Perceptions and Behavior Towards Clickbait Warnings, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2024.2323248

To link to this article: https://doi.org/10.1080/10447318.2024.2323248

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2024.2323248

"It is Luring You to Click on the Link With False Advertising" - Mental Models of Clickbait and Its Impact on User's Perceptions and Behavior Towards Clickbait Warnings

Ankit Shrestha, Arezou Behfar, and Mahdi Nasrullah Al-Ameen

Department of Computer Science, Utah State University, Logan, UT, USA

ABSTRACT

Clickbait, a social engineering attack performed through social media, tricks users through sensationalized or misleading posts into clicking on links that direct them to malicious websites. With the recent boom in social media, clickbait has become a substantial security concern, necessitating efforts from platforms and academia to control it. Despite these attempts, clickbait is effective due to the lack of users' knowledge. Therefore, we explore user mental models (thought processes about how something works) about clickbait to analyze their deficiencies and their influences on users' behavior towards clickbait warnings. To this end, we conducted an online study with 770 participants over MTurk to generate user mental models about clickbait and to evaluate the clickbait warnings conveying harm. Our findings suggest that a large portion of users have a simple mental model that fails to comprehend the dangers of clickbait, indicating the importance of warnings in supporting and educating users. Overall, our studies provide valuable insights into understanding the impact of clickbait mental models on users' online security behavior in social media and offer guidelines for future research in these directions.

KEYWORDS

Clickbait; mental models; warnings; quantitative study

1. Introduction

Social engineering attacks exploit humans, the weakest link in online security (Aldawood & Skinner, 2019; Khiralla, 2020). In fact, 90% of data breach incidents in the United States target the human elements through some form of social engineering.[1] Past incidents using social engineering attacks have resulted in severe consequences including sexual exploitation and extortion (Wittes et al., 2016). With the advent of social networking sites, public information about users, including their affiliated institutions, interests, and even their friends, are readily available (Huber et al., 2009; Krombholz et al., 2015; Sharevski et al., 2022). For instance, an attacker could use a photo of a social media user to create clickbait posts using Artificial Intelligence like Deepfakes and target his/her connections. Such availability of public information, therefore, adds another dimension to the effectiveness of social engineering attacks (Ajina et al., 2023; Allen et al., 2022; Hu & Apuke, 2023; Krombholz et al., 2015; Lewandowsky et al., 2012).

Social engineering attacks such as phishing have caused many problems but the virality of content and the lack of scrutiny in social media could result in severe damages to people and society (Allen et al., 2022; Geeng et al., 2020). In that regard, clickbait is a social engineering attack primarily carried out through social networking sites that use misleading or sensationalized headlines and images to trick users into clicking on malicious links (Avery et al., 2017; Li et al., 2022; Redmiles et al., 2018; Rides, 2017; Scott, 2021; Souza, 2015). With the increasing popularity of social networking sites, clickbait poses a substantial threat to the online safety of social media users (Aldawood & Skinner, 2019; Avery et al., 2017; O'Donnell, 2018; Redmiles et al., 2018; Rides, 2017). Clickbait is known to direct users to websites, including phishing sites and the sites spreading ransomware, viruses, Trojans, adware, and spyware (Avery et al., 2017; Redmiles et al., 2018; Rides, 2017; Souza, 2015). Clickbait also helps to spread misinformation (Zeng et al., 2020), which can threaten public health (Bin Naeem & Kamel Boulos, 2021; Javed et al., 2020; Pine et al., 2021; Xiang et al., 2023; Zhang et al., 2022) and safety (Faris et al., 2017; Marwick & Lewis, 2017; Peck, 2020; Sylvia Chou et al., 2020; Tasnim et al., 2020; Vasudeva & Barkdull, 2020). This is further aggravated by social media users' lack of knowledge and awareness about clickbait, which situates them in a vulnerable position (Huang et al., 2015; Urakami et al., 2022).

While social engineering attacks account for 98% of cybersecurity incidents,[2] the human aspect of the attacks is rarely studied. Therefore, in our work, we focused on answering "how users understand clickbait" and" why users interact with it." Here, our first step was to understand the users' existing concepts about clickbait and the gaps within their understanding. To that end, mental models[3] that represent the user's understanding of a concept provided us with

CONTACT Ankit Shrestha ankit.shrestha@usu.edu Department of Computer Science, Utah State University, Logan, UT, USA

a viable method to group users with similar knowledge together (Johnston-Laird, 1983; Young, 2008). While user groups could also be formed based on demographics (age, sex, location), grouping through mental models allowed us to understand the social media users' behavior towards clickbait based on their existing ideas and perceptions (Kaptein et al., 2015; Liu et al., 2016). Further, grouping users through mental models helps with the generalization and ideation of personalized solutions for these groups (Kaptein et al., 2015; Liu et al., 2016). Using mental model based user groups also helped us unveil how the understanding of clickbait impacts users' perceptions and behavior toward mitigation attempts against it.

In fact, social media platforms (Babu et al., 2017; Gleicher, 2019; Roth & Harvey, 2018; Safety, 2019) and academia (Bhuiyan et al., 2021; Hassan et al., 2019; Lewandowsky et al., 2012; Schul, 1993) pushed forward several attempts to counter clickbait. These attempts mostly included detection (Agrawal, 2016; Chien et al., 2022; Karande et al., 2021; Zheng et al., 2018; Zhou, 2017) and moderation (Babu et al., 2017; Gleicher, 2019; Roth & Harvey, 2018; Safety, 2019) which have limitations (D. Molina et al., 2021; Karande et al., 2021). Only a few works focused on supporting users through interventions to make informed decisions against clickbait (Bhuiyan et al., 2018; Chakraborty et al., 2016; Hassan et al., 2019). However, these works only classified posts as clickbait, failing to increase users' awareness and knowledge. Therefore, clickbait remains effective in social networking sites, aggravated by users' unwillingness to investigate low-credibility posts (Allen et al., 2022; Geeng et al., 2020). We addressed these gaps in our study.

In that regard, we first designed warnings conveying the harm of clickbait, one of the most effective techniques of changing user behavior (Abraham & Michie, 2008; Michie et al., 2013). The choice of conveying consequences was further motivated by the invisible nature of consequences from clickbait, giving users a fake sense of security (Aldawood & Skinner, 2019). For instance, users rarely know their information is stolen using cookies on sites that clickbait leads to. Being unaware of such consequences habituates the users to clickbait and creates a conception that clickbait is often harmless. In our warnings, the consequences of clicking on clickbait were further delivered in two variations - logical listing of information (Logical Warning) (Amgoud et al., 2007) and emotional story with characters (Emotional Warning) (Lan et al., 2022).

Based on the gap in the literature and the designed warnings, we investigated the following research questions:

RQ1: What are the different users' mental models of clickbait?
RQ2: How do the mental models of users influence their behavior towards clickbait and the warnings designed against it?

To address these questions, we conducted a study with 770 participants on Amazon Mechanical Turk (Mturk). In this study, we asked them to interact with and evaluate a clickbait post to understand the users' perception and behavior towards clickbait. Then, we asked them about their understanding of clickbait and derived six mental models from our analysis (RQ1). Next, we asked them to interact and evaluate the designed Logical and Emotional warnings against clickbait. The findings from the online study informed us about the perceptions and behaviors of users with different mental models towards clickbait (RQ2). We observed that most mental model groups lacked comprehension of the dangers associated with clickbait interactions, rendering them vulnerable to such attacks. Our findings also unveiled the preferences and behavior of these groups towards the designed Logical and Emotional warnings (RQ2).

Here, we acknowledge that mental models, a representation of users' understanding of a concept, are diverse and may even be non-exhaustive. However, our goal was not to exhaust all possible mental models of clickbait. Instead, our contributions include the knowledge about users' understanding of clickbait and how these understandings can shape their online security behavior. In doing that, our findings provide valuable insights into users' mental models of clickbait and its influence on their perceptions and behavior regarding clickbait and warnings against it. Finally, these findings point to a set of recommendations, including mental model augmentation and personalization of interventions.

2. Related work

We discuss prior works that help us understand the efficacy of clickbait in tricking users in §2.1 and our motivation to unveil their understanding of clickbait in §2.2. Then, we discuss the prior mitigation attempts against clickbait, leading us to design our interventions in §2.3. We present the evaluation of these interventions in our results through the sense-making lenses of the users' mental models of clickbait.

2.1. Importance of safeguarding users against clickbait

In social media and online settings, the most significant security threats are posed by social engineering attacks (Hadnagy, 2010; Indrajit, 2017; Kee & Deterding, 2008). There are real-life precedents where social engineering attacks resulted in severe consequences. For instance, Wittes et al. (2016) reported the sexual exploitation of roughly 230 people from a single attacker where users were tricked into downloading malware. Clickbait, a social engineering attack primarily performed through social media, directs users to malicious sites, including those spreading ransomware, viruses, Trojans, adware, and spyware (Avery et al., 2017; O'Donnell, 2018; Redmiles et al., 2018; Rides, 2017). Even worse, clickbait helps spread misinformation that severely impacts public health and safety (Vasudeva & Barkdull, 2020; Zeng et al., 2020; Zhang et al., 2022).

While we acknowledge several attempts from social media platforms to limit clickbait, users still encounter it

regularly, attributing to its effectiveness (Gleicher, 2019; Roth & Harvey, 2018). That begs the question, "Why is clickbait effective?" Literature provides us with three reasons explaining the users' inclination towards interaction with clickbait. First, clickbait increases its effectiveness by creating a curiosity/information gap where users feel rewarded with answers when they click on it (Li et al., 2022; Scott, 2021). It relates to the cognitive principle of relevance, which explains that users seek to maximize the relevance of the information (Clark, 2013; Sperber et al., 1995; Sperber & Wilson, 1986). Second, users lack education and awareness to identify clickbait and therefore do not understand the importance of avoiding it (Huang et al., 2015; Urakami et al., 2022). Third, the influence of clickbait is further aggravated when it aligns with the beliefs of the users (Allen et al., 2022; Wineburg & McGrew, 2019). It can be explained by the communicative principle of relevance that states relevance is optimal when it accounts for user preferences and abilities (Clark, 2013; Sperber et al., 1995; Sperber & Wilson, 1986). However, few studies have focused on understanding users' perceptions of clickbait and countermeasures against it. We address this gap in our work (RQ1 and RQ2).

2.2. Motivation to understand users' mental models

Understanding user mental models can reveal existing gaps in their knowledge and inform how interventions may be designed for them. In that regard, Norman (2013) explains the concepts of system image and mental model. System image is the information about the system available to the users (e.g., what users can understand from a clickbait) (Norman, 2013). A mental model is what users understand from the system (in this case, clickbait) (Norman, 2013, 2014). While Norman (2013) suggests system images to be elaborate so that users can understand the designer's concepts and intentions, attackers aim to obscure them in clickbait, resulting in incorrect and incomplete mental models. For instance, consequences of clickbait are not readily visible, leading users to perceive them as harmless (Vance et al., 2017). Users rarely realize that attackers steal their information through cookies or feed them misinformation when they click on clickbait. To that end, mental models help us make sense of users understanding of clickbait and provide us with the necessary background to inform future designs of interventions against it.

Several works have suggested that contextualizing information in interventions based on users can enhance the understanding of a concept (Kaptein et al., 2015; Liu et al., 2016). However, such contextualization based on sex, age, and location may be impractical due to the differences in their understanding. To that end, mental models provide a practical choice to effectively contextualize interventions against clickbait based on the user's existing understanding (Johnston-Laird, 1983; Young, 2008). Several studies have worked on identifying mental models about concepts such as the Internet and security tools (Dumaru et al., 2023; Kang et al., 2015; Oates et al., 2018; Paudel et al., 2023; Wu & Zappala, 2018). The study of Thatcher and Greyling (1998) depicted a hierarchical categorization based on the complexity of users' mental models of the Internet. On the other hand, Kang et al. (2015) presented a binary categorization of simple vs. articulate mental model. In another study, Wu and Zappala (2018) identified mental models to understand how users perceive the working of encryption. Oates et al. (2018) revealed users' mental models of privacy from the illustrations created by users about what privacy means to them. In a separate study, Abu-Salma and Livshits (2020) evaluated the user interface of the private mode in different browsers, revealing that the existing browser disclosures fail to illustrate the primary objectives of private browsing to users. However, none of these studies explored the mental models of clickbait. To our knowledge, our study is the first one to do that. Further, the issues relating to security become prominent when there is a gap in users' understanding of concepts (mental models). Due to these reasons, we first focus on understanding users' mental models of clickbait.

2.3. Clickbait mitigation: Conveying consequences using logic and emotion

While understanding mental models of clickbait is essential, we also focused on understanding the ideas and interventions that may work for the different mental models. To that end, much of the existing literature on clickbait focuses on detecting and moderating content (Chakraborty et al., 2016; Chien et al., 2022; Geeng et al., 2020; Gleicher, 2019; Heuer & Glassman, 2022; Roth & Harvey, 2018; Safety, 2019). However, moderation can lead to problems. The existing clickbait detection methods might not be reliable (D. Molina et al., 2021) – the best method using state-of-the-art language models still have 4.68% error (Karande et al., 2021). In such a situation, completely blocking posts will lead to blocking a substantial number of non-clickbait posts and vice versa. Therefore, we shift our focus towards supporting users through interventions while allowing users to make informed decisions.

In that regard, only a few works focused on supporting users to make informed decisions about clickbait and misinformation (Bhuiyan et al., 2018, 2021; Chakraborty et al., 2016; Ecker et al., 2010; Hassan et al., 2019; Konstantinou et al., 2024; Lewandowsky et al., 2012; Schul, 1993). These works primarily focused on identifying clickbait but created interventions that only classified posts as clickbait for the users (Chakraborty et al., 2016). However, studies show that clickbait is effective (see §2.1) despite these efforts due to users' lack of knowledge (Huang et al., 2015; Urakami et al., 2022) and their unwillingness to investigate low-credibility posts (Allen et al., 2022; Geeng et al., 2020). Therefore, we focus on informative interventions to support users. In that regard, we primarily focus on interventions conveying the consequences of clickbait. Our selection is based on prior studies that show conveying consequences is one of the most effective techniques to change user behavior (Abraham & Michie, 2008; Kaiser et al., 2021; Michie et al., 2013).

In conveying the consequence of clickbait, we focus on two approaches- Logical and Emotional. Logical and Emotional warnings have been extensively used with some degree of success in multiple fields (Amgoud et al., 2007; Fillenbaum, 1976; Lan et al., 2022; Shrestha et al., 2023). Moreover, our selection of these approaches stems from prior works highlighting logic and emotion as the two most effective methods of delivering information to persuade change in user behavior (Cronkhite, 1964; Woolbert, 1918). However, such approaches are yet to be explored against clickbait. In that regard, our study aims to understand how users with different mental models perceive these two approaches (RQ2), and based on our findings, we recommended scopes of personalization of warnings for different mental models.

3. Methodology

3.1. Intervention design

We conveyed the consequences of clickbait using two approaches- logic and emotion. In the Logical warning, we listed negative consequences with text and graphics (Figure 1(a)). Using multiple modes of information aligns with dual-code theory, improving the effectiveness of the conveyed information (Mayer, 2014; Moreno & Valdez, 2005).

In our emotional warning, we created a story due to its efficacy in persuasion (Dessart & Standaert, 2023; Murnane et al., 2020; Simmons, 2019). In the story, a character clicked on the post and faced the consequences. Here, we used emotional expressions in the faces of the characters to appeal to the users' emotions (Figure 1(b)). The story is depicted through graphics as they convey information more effectively (Kumaraguru et al., 2010; Paivio, 2006).

3.2. Online study

We collected the perceptions of clickbait (to generate mental models) (RQ1) and evaluated the interventions (RQ2) through an online survey over MTurk. The survey was created using JavaScript for frontend and node.js for backend. Participants had to be at least 18 years old and live in the United States or Canada to participate in our study. On average, participants took approximately 15 minutes to complete the study. We compensated them with USD 2.5. The Institutional Review Board approved the study at our university (IRB #12735).

3.2.1. Survey questionnaire

In our study, we used four parameters to rate the clickbait post: Interest, Likelihood (to click), Safety (of the link), and Knowledge (about the content the post leads to) (Table 1). To rate the clickbait warnings, we used four parameters from the User Experience Questionnaire (UEQ)[4]: Attractiveness, Perspicuity, Efficiency, and Dependability. The UEQ measures an artifact's pragmatic (practical) and hedonic (relating to pleasure) qualities. These four parameters were included as we focus on understanding the pragmatic qualities of the interventions (RQ2). We also added custom questions to measure the Effectiveness, Satisfaction, Informativeness, and Likelihood since we did not find validated scales (Table 2). All the survey questions used a 7-point Likert scale (−3 to 3). Additionally, we used six attention-check questions presented to the participants in random order, following the guideline from prior work (Ipeirotis et al., 2010; Kung et al., 2018).

3.2.2. Procedure

Upon agreeing to the Informed Consent Document (ICD), participants were provided an overview of our study (Figure 2). Initially, participants were asked to interact with a mock clickbait post (without informing them that the post is clickbait) and evaluate the post on four parameters- Interest, Likelihood, Safety, and Knowledge (Table 1). Then, participants were asked about their understanding of clickbait using an open-ended question, "What do you

Table 1. Survey questions used to rate the clickbait post.

Parameter	Questions
Interest	The post is interesting
	The post is attractive
Likelihood	I am likely to click on this post
Knowledge	I already know what is inside the post
Safety	The post is safe to click on

(a) Logical Warning

(b) Emotional Warning

Figure 1. The two warning variations used in the online study.

understand by the term Clickbait?" (Note: The open-ended question is used to derive the mental models of clickbait). We then explained to them what we meant by the term "clickbait" in the context of our study. Participants then interacted with the two variations of the warnings: Logical and Emotional. After each variation, they responded to Likert-scale questions mentioned in §3.2.1 and open-ended questions about what they liked and disliked in the warnings. At the end, participants answered a set of demographic questions.

3.2.3. Quality control

While MTurk workers may not always pay close attention to the instructions (Oppenheimer et al., 2009), we followed the guidelines from prior studies (Kung et al., 2018; Peer et al., 2014) to increase the quality of responses. Our survey required that the participants had above a 95% HIT approval rate.[5] Moreover, we included six attention check questions in our survey and only included 770 responses that correctly answered all six of our attention check questions. Further, our analysis of the answers to the question, "What do you understand by the term Clickbait?" revealed 48 responses that were removed from our analysis due to three reasons – (1) lack of understandability (For instance, "Clickbait is the important is a post."), (2) extreme shortness (For instance, "clickbait"), and (3) irrelevance (For instance, "I think it is a TV show").

Table 2. Custom survey questions used to rate the clickbait warnings (UEQ questions are available in footnote 4).

Parameter	Questions
Effectiveness	The warning motivated me to avoid the post
	After seeing this warning, I am likely to avoid clickbait even without any warning
	in the future
	The information in the warning helped me to make an informed decision
	I am likely to adopt this warning in real life
Satisfaction	I am satisfied with the time required to interact with the warning
	I am satisfied with the effort required to interact with the warning
	The interaction with the warning was exhausting/frustrating
	I am satisfied with the information in the warning
Informativeness	The information in the warning is accurate
	The information in the warning is consistent
	The information in the warning meets your expectations
	The consequences of ignoring the warning were clear
Likelihood	I am likely to click on the post despite seeing the warning

3.2.4. Analysis

We include the remaining 722 participants after quality checks in our analysis (Baxter et al., 2015; Boyatzis, 1998; Braun & Clarke, 2006). We first performed inductive thematic analysis on the responses about users' understanding of "Clickbait." To that end, two independent researchers coded each response, developed codes, and assigned a mental model. The inter-coder reliability in the thematic analysis was 88.78%. We report our findings based on the users' mental models.

Using a similar approach, we conducted a thematic analysis of the responses relating to the likes and dislikes of the warnings. The inter-coder reliability was 88.6% in this case.

We used statistical tests to analyze our quantitative results. We consider results significant when we find $p < .05$. When comparing two conditions, we use a Wilcoxon signed rank test for the matched pairs of subjects since the study was within subjects and the data distribution was not normal. Wilcoxon tests are similar to t-tests but do not assume the distributions of the compared samples, which is appropriate for our collected data.

Our analysis of the warnings was conducted through the lenses of the users' mental models, where we compared the two warning variations based on user groups created

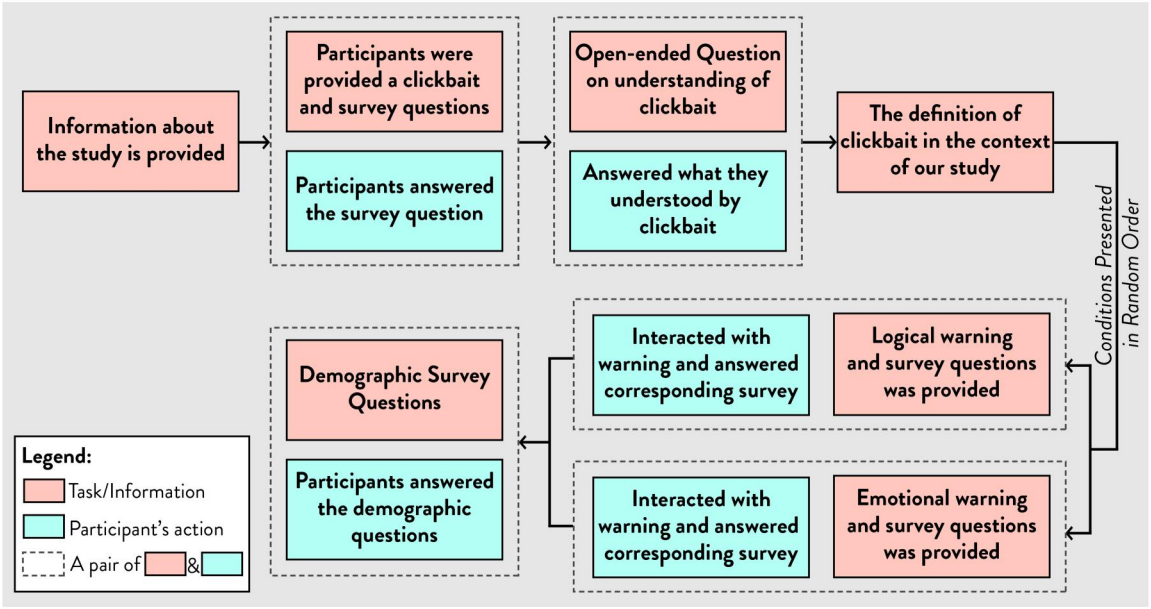


Figure 2. Flow of the survey in the online study.

through their mental models. Our goal with such an analysis is to unveil the perspectives of the two warnings of particular user groups and understand which approach (logical or emotional) may be more suitable for these groups.

3.2.5. Demographics

Table 3 presents the demographic summary of the participants from our online study.

4. Results

We first extracted the users' mental models of clickbait based on our analysis. We observed that users' mental models overlap partially, whereas new ideas may be added to this partial overlap. These partially overlapping mental models made creating groups of similar users difficult. To overcome that challenge, we conducted a mental model decomposition (Figure 3) by extracting simpler unitary concepts representing a single aspect of understanding. Here, we found that users make sense of clickbait in two broad ways: (1) how it works (see §4.1), and (2) what it aims to achieve (see §4.2). The decomposition of mental models helped us to understand the users' perceptions at a more granular level. Our in-depth analysis revealed a set of mental models under each of these two sensemaking categories. We further found instances where users make sense of clickbait based on its working and goal together (see §4.3).

We structure the findings section based on the mental models identified from our research. We first present a mental model and then follow it with the evaluation of clickbait and warnings against it by participants with such a mental model. Since there are multiple mental models presented in the paper, we wanted to make sure readers understand how each of these participant groups evaluated and behaved towards clickbait and warnings against it. Therefore, we structure the findings so the readers can understand one mental model group and their perceptions and behavior before moving on to the next one. In answering RQ1, the mental models are explained in the first paragraph of each section that follows: §4.1, §4.1.1, §4.1.2, §4.1.3, §4.2, §4.2.1, §4.2.2, §4.2.3, and §4.3. In answering RQ2, the perception and behavior towards clickbait and warnings against it are presented in the subsequent paragraphs in these sections.

Table 3. Demographic information of the participants in the online study (*N*=number of participants).

Demographic	Demographic group	*N*
Gender	Male	411
	Female	299
	Other	7
	I prefer not to answer	5
Age range	18–24 Years old	27
	25–29 Years old	83
	30–34 Years old	144
	35–39 Years old	140
	40–44 Years old	113
	45–49 Years old	59
	50–54 Years old	46
	55–59 Years old	52
	60–64 Years old	38
	Above 65 years old	15
	I prefer not to answer	5
Race	White	558
	Asian	50
	Black/African American	39
	Hispanic or Latino	21
	Native American	5
	Mixed Race	38
	Other	2
	I prefer not to answer	9
Education	Less than high school	6
	High school graduate	147
	Two-year college degree	104
	Four-year college degree	346
	Graduate degree (MS/Ph.D.)	103
	I prefer not to answer	7
	Other	9

4.1. Working mechanisms of clickbait (shortened as Work MM)

We observed that 31.85% of participants defined clickbait based on its working without mentioning its goals. These participants tried to make sense of clickbait by explaining how it enticed users but did not relate it to why it is used. For instance, one participant commented,

> Clickbait is where the headline says something interesting that makes you want to click on it. Then once you do, the story is either not what was advertised or it is a long story about something else with only the slightest amount of relevance to what you thought it would be.

Participants with *Work MM* have moderately high interest and likelihood to click on the post (Figure 4). Further, the participants perceive the post to be safe, indicating that most of these participants may be unable to identify the

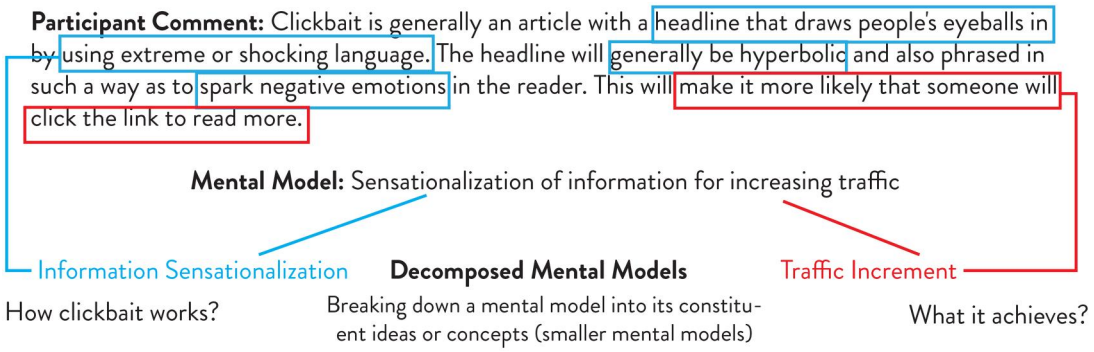


Figure 3. Decomposition of mental models into constituent concepts.

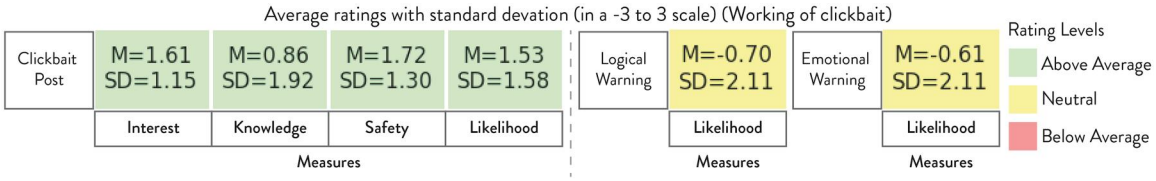
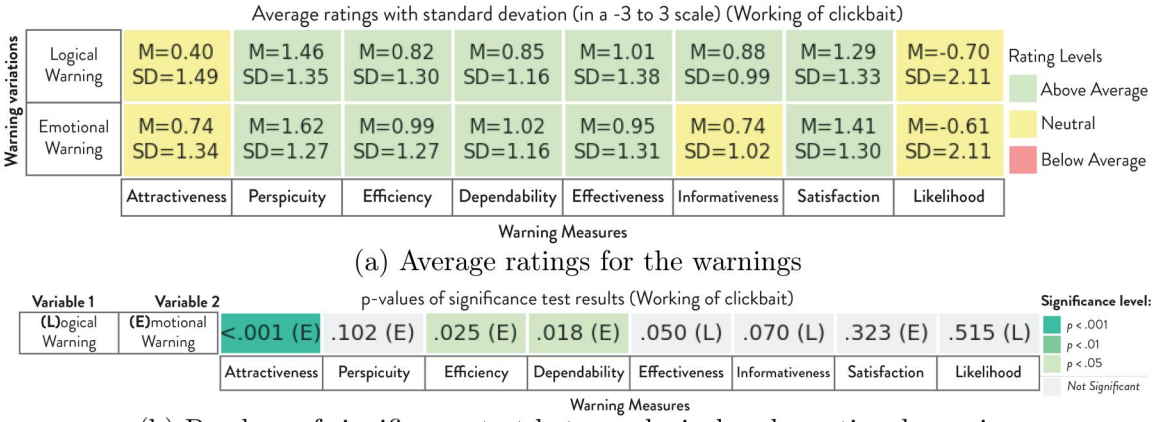


Figure 4. Average ratings for the clickbait post along with the likelihood to click with the warning (working of clickbait mental model).



(a) Average ratings for the warnings

(b) P-values of significance test between logical and emotional warning

Figure 5. Working mechanism of clickbait mental model.

post as clickbait. However, the two warning variations decreased the likelihood of clicking on the post for *Work MM* participants by more than two points (Figure 4). When comparing the likelihood of the participants to click on the post with and without the warnings, we observed that both logical ($W = 852.5$, $p < .001$) and emotional ($W = 1128.0$, $p < .001$) warning significantly reduce the participant's likelihood indicating that these warnings can be helpful against clickbait.

Upon diving deeper, participants who understand clickbait based on its working mechanisms found these warnings above average in most ratings (Figure 5(a)). They found the logical warning above average in terms of perspicuity, efficiency, dependability, effectiveness, informativeness, and satisfaction, which could explain the significant reduction in the likelihood of clicking on the post. Some participants with *Work MM* found the logical warning straightforward, explaining above-average ratings for informativeness, perspicuity, and efficiency. One participant said,

> I think implementing the color red [in the overlay] is a wise move. Additionally, I also appreciate the simplicity. Sometimes less is more.

Similarly, the *Work MM* participants found the emotional warning above average in perspicuity, efficiency, dependability, effectiveness, and satisfaction (Figure 5(a)). Open-ended responses support our results as some participants with *Work MM* found the story conveyed in the warning meaningful and transparent. One participant said,

> I really liked the image of the woman on the laptop. It was very clear what this was trying to convey. The keep scrolling button was prominent and a different color, which made it stand out more than the warning and the option to proceed. I also like that the warning has a red background to grab my attention.

The *Work MM* participants found the emotional warning significantly more attractive, dependable, and efficient than the logical one (Figure 5(b)). Open-ended responses show that some participants found the depiction of the emotional story to be friendly and pleasant. One participant said,

> I liked the character animation because they are colorful and attention-getting.

The emotional stories further highlight the information delivered through the characters that users can relate to, which could be perceived as more reliable.

Upon diving deeper, we identified three decomposed mental models under the sensemaking lens of how clickbait works. We discuss each of these mental models in more detail in §4.1.1, §4.1.2, and §4.1.3.

4.1.1. Information Sensationalization (shortened as Sensation MM)

Information Sensationalization is the mental model where participants believed clickbait worked by exaggerating either the thumbnail or the headline. 52.63% of participants had such a mental model. That may be standalone or in combination with other mental models. For instance, one participant commented,

> Clickbait is a teaser post that entices you to click on the link. Once you click on the link, you are most likely going to be disappointed in the news story. It usually does not live up to the tease.

Participants with *Sensation MM* are quite interested and likely to click on clickbait without any warning (Figure 6). They also think the post is safe to click on, as evident by the high safety score. However, the likelihood scores

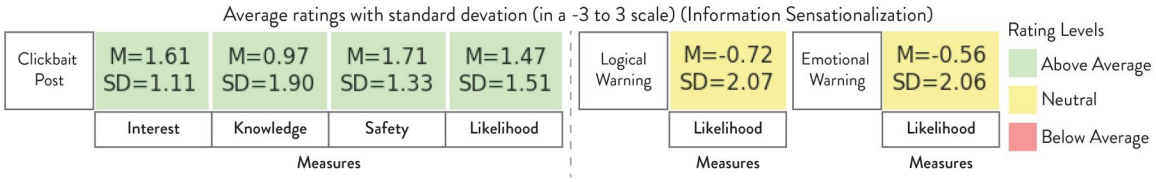
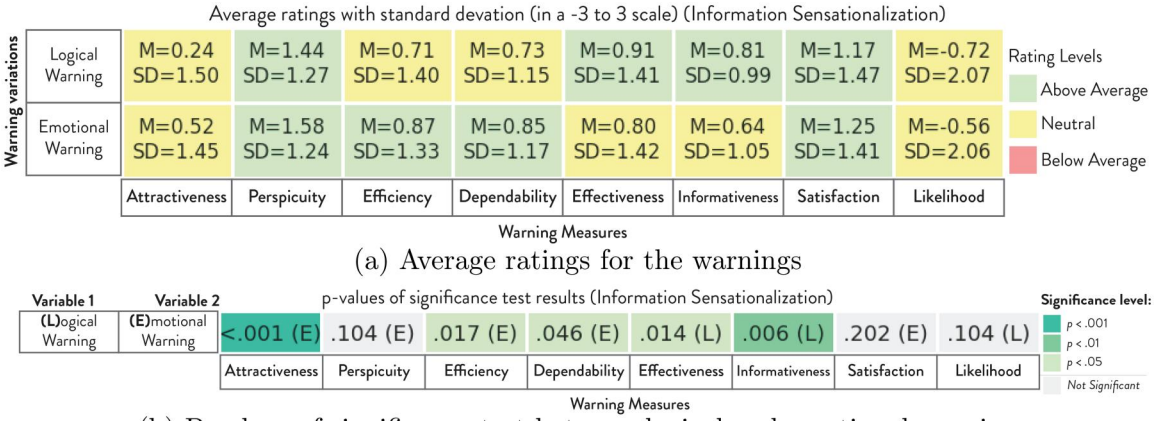


Figure 6. Average ratings for the clickbait post along with the likelihood to click with the warning (information sensationalization mental model).



(a) Average ratings for the warnings

(b) P-values of significance test between logical and emotional warning

Figure 7. Information sensationalization mental model.

decreased considerably with the warnings, highlighting its importance (Figure 6). The significance test between the likelihood of clicking with and without the warnings revealed that both the logical ($W = 3249.0$, $p < .001$) and emotional ($W = 4250.5$, $p < .001$) warning significantly reduced the participant's likelihood to click on clickbait.

Sensation MM participants found the logical warning above average in perspicuity, effectiveness, informativeness, and satisfaction (Figure 7(a)). Similarly, they rated the emotional warning above average in perspicuity, efficiency, dependability, and satisfaction (Figure 7(a)).

These participants found the emotional warning significantly more dependable and efficient (Figure 7(b)). Some participants found the characters in the story relatable, which increased their trust in the information provided by these characters. One of them commented,

> I liked that it was very personable and kind of fun with the characters/graphics. It is more "warm" and interactive than just a more mechanical/data-driven warning.

They also found the logical warning significantly more effective and informative (Figure 7(b)). They perceived the logical information in the warning as more valuable than a story. Such a perception could also have led to increased effectiveness of the warning. One participant commented,

> I liked the explanations that were given as to why you should not click on the link. They help to inform you of the possible dangers that may be present.

4.1.2. Deceptive Presentation (shortened as Deception MM)

The understanding of clickbait, where participants believed in the involvement of some trickery, lying, or non-factual information, was termed as *Deceptive Presentation*. 64.95% of participants had such a mental model (standalone or in combination). For instance, one participant commented about the non-accurate information used in clickbait,

> It refers to a headline or a picture that is not accurate to the actual content of the link. It is luring you to click on the link with false advertising.

Participants with *Deception MM* have moderately high interest and likelihood to click on the post (Figure 8). These participants perceived the post as safe to click on. However, with the warnings, the likelihood of clicking on the post decreased by more than two points (Figure 8). Further, we observed that both logical ($W = 4148.0$, $p < .001$) and emotional ($W = 5293.5$, $p < .001$) warnings significantly reduced the participant's likelihood to click on clickbait.

Participants with *Deception MM* found the logical warning above average in perspicuity, dependability, effectiveness, informativeness, and satisfaction (Figure 9(a)). Similarly, they rated the emotional one above average in perspicuity, efficiency, dependability, effectiveness, and satisfaction (Figure 9(a)).

The emotional warning was rated significantly more dependable, efficient, understandable, and satisfactory than the logical one (Figure 9(b)). Since these participants were aware of the deception of clickbait, instead of providing just logical information, the stories showing characters facing the consequences could be more relatable, resulting in better perspicuity, dependability, and satisfaction. One participant said,

> I like the information given in the clouds above the figureheads; it helps me to visualize what might happen to me if the link is clicked.

They also found the logical warning significantly more effective and informative than the emotional one

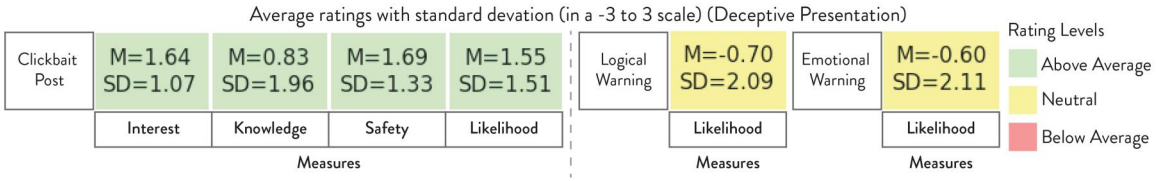
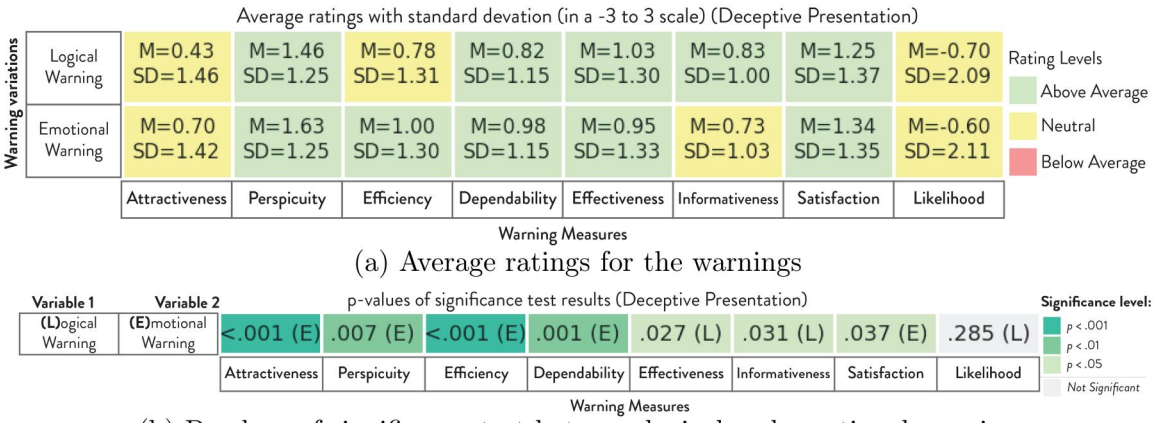


Figure 8. Average ratings for the clickbait post along with the likelihood to click with the warning (deceptive presentation mental model).



(a) Average ratings for the warnings

(b) P-values of significance test between logical and emotional warning

Figure 9. Deceptive presentation mental model.

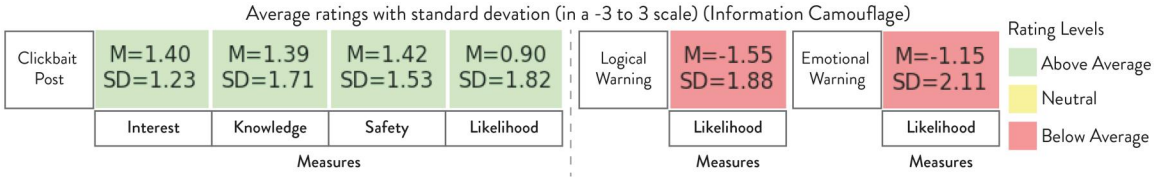


Figure 10. Average ratings for the clickbait post along with the likelihood to click with the warning (information camouflage mental model).

(Figure 9(b)). Like the *Sensation MM* participants, *Deception MM* participants also found the logical information valuable in informing their decision to avoid clickbait. One participant said,

> I like that it is in your face; the information is right there and easy enough to read and understand.

4.1.3. Information Camouflage (shortened as Camouflage MM)

The mental model where participants thought clickbait works by hiding the most critical information from the *p* was termed as *Information Camouflage*. 9.14% of the participants had this mental model (standalone or in combination). For instance, one participant commented,

> Clickbait is when an ad words things in a way that makes you want to click to see more. So they initially give you little information so that you want to click on the ad to find out more information. And also so that maybe they can get a commission off your click.

Participants with *Camouflage MM* have comparatively lower interest and higher knowledge about the post. They are also comparatively less likely to click on the post (Figure 10) as they understood that clickbait is trying to hide information. Further, these participants have lower scores for the perceived safety of the post, indicating that a larger portion of them may be aware that the post is clickbait and can harm them. With the warnings, the likelihood scores decreased even further (Figure 10). In fact, both logical ($W = 110.0$, $p < .001$) and emotional ($W = 118.0$, $p < .001$) warnings significantly reduced the participant's likelihood to click on clickbait.

Camouflage MM participants found the logical warning above average in perspicuity, effectiveness, informativeness, and satisfaction (Figure 11(a)). Similarly, they rated the emotional one above average in perspicuity, effectiveness, and satisfaction (Figure 11(a)). We observed that the warning ratings are comparatively lower for these participants than those with a different mental model. Since these participants were already less likely to click on the post without any warnings, they might perceive the warnings are not very useful or practical in their cases. We can infer this from the low likelihood of clicking on the post after seeing the warnings despite poor scores across other parameters measuring user experience.

When comparing the two warning variations, these participants found the logical warning significantly more informative than the emotional one (Figure 11(b)). Since they know the hiding of information in clickbait, they may have disliked depicting information through a story instead of a logical list of information (a roundabout way instead of

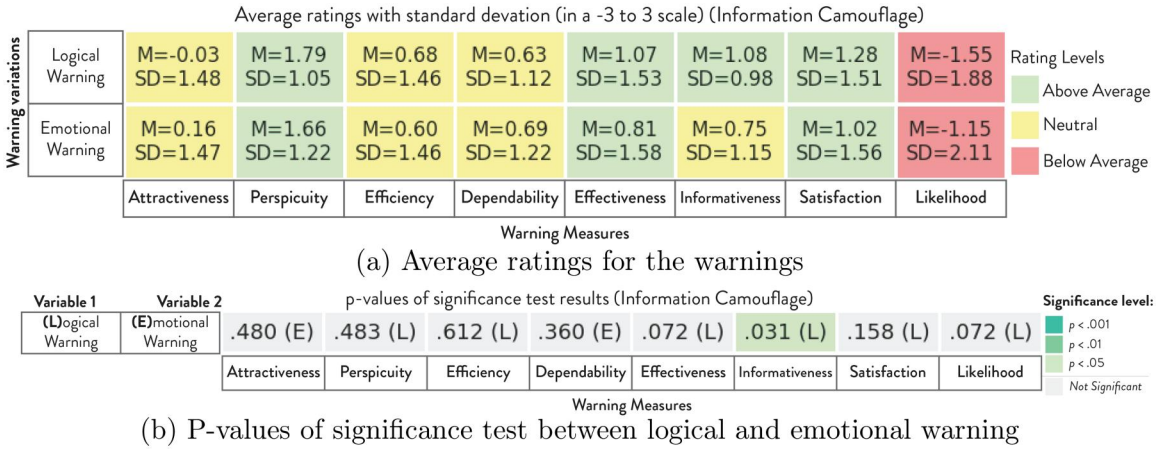


Figure 11. Information camouflage mental model.

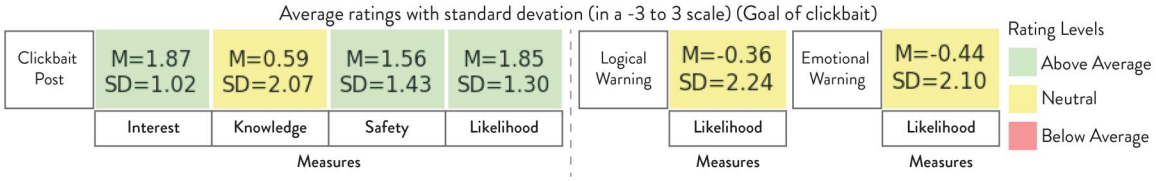


Figure 12. Average ratings for the clickbait post along with the likelihood to click with the warning (goal of clickbait mental model).

a direct listing). One participant describing the logical warning said,

> I like that it clearly describes all the bad points of clickbait articles.

4.2. Goal of clickbait (shortened as Goal MM)

We observed that 10.80% of participants defined clickbait based on its goals without mentioning its working. These participants tried to make sense of clickbait on why it is created but did not relate it to how it may work. For instance, one participant commented,

> In my mind, clickbait's purpose is to spread lots of the advertisements through the websites. Then. steal user information.

Participants with *Goal MM* have comparatively higher interest and a higher likelihood to click on the post even though they have a lower perception of safety (Figure 12). Further, we see a lower knowledge about the post among these participants. Here, logical and emotional warnings decreased the likelihood of clicking on the post for by more than two points (Figure 12). Moreover, both logical ($W = 90.0$, $p < .001$) and emotional ($W = 101.0$, $p < .001$) warnings significantly reduced the participants' likelihood to click on clickbait.

These participants rated both warnings above average in most parameters (Figure 13(a)). The logical warning was above average in perspicuity, efficiency, dependability, effectiveness, informativeness, and satisfaction. Similarly, the emotional one was rated above average in attractiveness, perspicuity, efficiency, dependability, effectiveness, informativeness, and satisfaction. There was no significant difference between the ratings for two warnings (Figure 13(b)).

We identified three mental models under the sensemaking lens of what clickbait aims to achieve. We discuss each of these mental models in more detail in §4.2.1, §4.2.2, and §4.2.3.

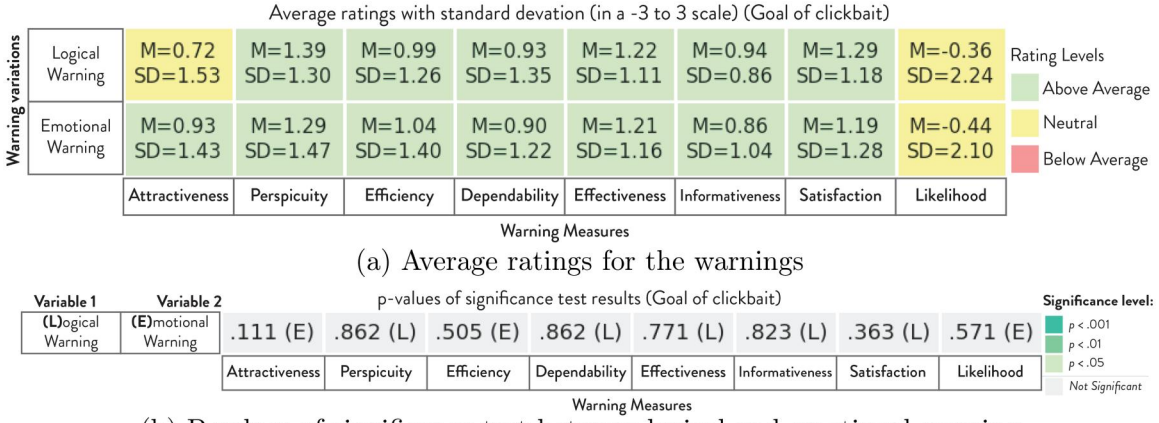
4.2.1. Traffic Increment (shortened as Traffic MM)

The mental model where participants thought clickbait was created to get more users to visit a website to generate traffic was termed as *Traffic Increment*. 58.72% of the participants had this mental model. For instance, one participant commented,

> Clickbait is a form of content designed to gather clicks on the search engine result pages. With clickbait, companies attempt to generate traffic on their blogs or websites.

Participants with *Traffic MM* have high interest, low knowledge about the post, and a moderately high likelihood to click on the post (Figure 14). They also consider the post safe, as evidenced by their high score for perceived safety. Such scores could result from these participants considering clickbait to be simply a tool to get traffic to the website. The results indicate that these users must be informed about the risks associated with clicking on the post. When the warnings informed the participants, the likelihood scores decreased by more than two points (Figure 14). Here, both logical ($W = 3091.5$, $p < .001$) and emotional ($W = 3331.5$, $p < .001$) warnings significantly reduced the participant's likelihood to click on the post.

Traffic MM participants found the logical warning above average in perspicuity, efficiency, dependability, effectiveness, informativeness, and satisfaction (Figure 15(a)). Similarly, they rated the emotional one above average in perspicuity, efficiency, dependability, effectiveness, and



(a) Average ratings for the warnings

(b) P-values of significance test between logical and emotional warning

Figure 13. Goal of clickbait mental model.

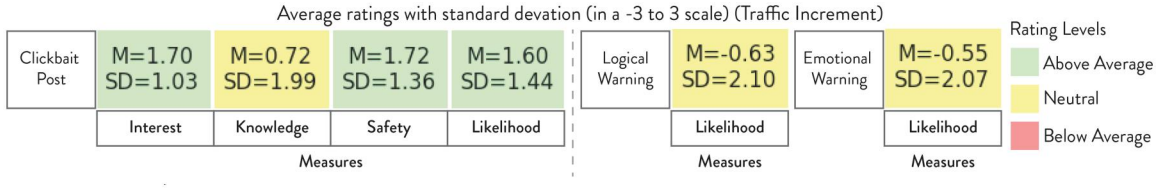
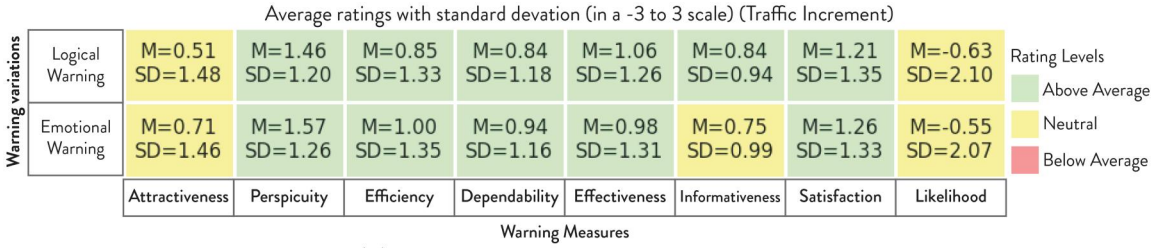
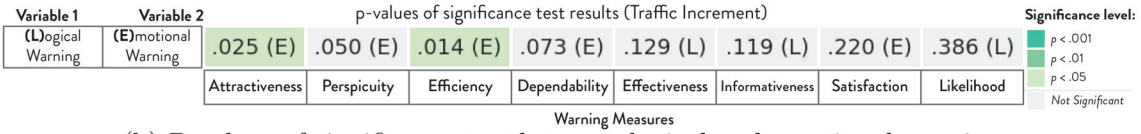


Figure 14. Average ratings for the clickbait post along with the likelihood to click with the warning (Traffic increment mental model).



(a) Average ratings for the warnings



(b) P-values of significance test between logical and emotional warning

Figure 15. Traffic increment mental model.

satisfaction (Figure 15(a)). These participants also found the emotional warning significantly more attractive and efficient than the logical one (Figure 15(b)). One participant mentioned,

> The information is presented clearly and effectively in a visually appealing way, making it very easy to understand what may happen if were to proceed anyway.

4.2.2. Financial Benefit (shortened as Financial MM)

Participants with *Financial Benefit* mental model thought that clickbait is a tool for advertisement and generating income. 13.69% of the participants had such a mental model. For instance, one participant commented,

> It is when you have some enticing content in the title that makes you want to click on the article. The intent of the creator is to get you on their page that is full of ads so they can gain views and make money. The article will be split up into different pages or very long on one page with tons of ads to scroll through.

Participants with *Financial MM* have high interest but comparatively lower likelihood to click on the post (Figure 16). That may be due to participants' higher knowledge about the post content, as observed in the ratings. With the warnings, the likelihood scores decreased by more than two points (Figure 16). The significance test revealed that both logical ($W = 306.0$, $p < .001$) and emotional ($W = 308.5$, $p < .001$) warnings reduced the participant's likelihood to click on clickbait.

These participants found the logical warning above average in perspicuity, effectiveness, informativeness, and satisfaction (Figure 17(a)). Similarly, they rated the emotional one above average in perspicuity, efficiency, dependability, effectiveness, informativeness, and satisfaction (Figure 17(a)). Participants

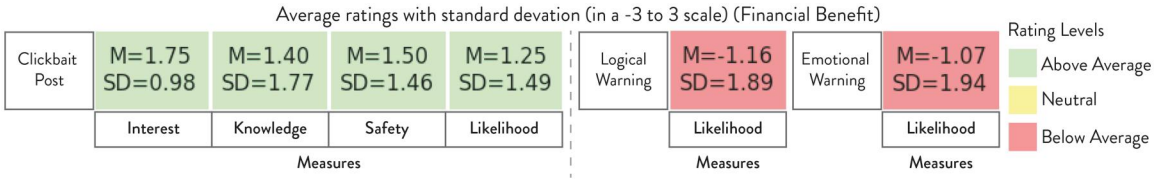
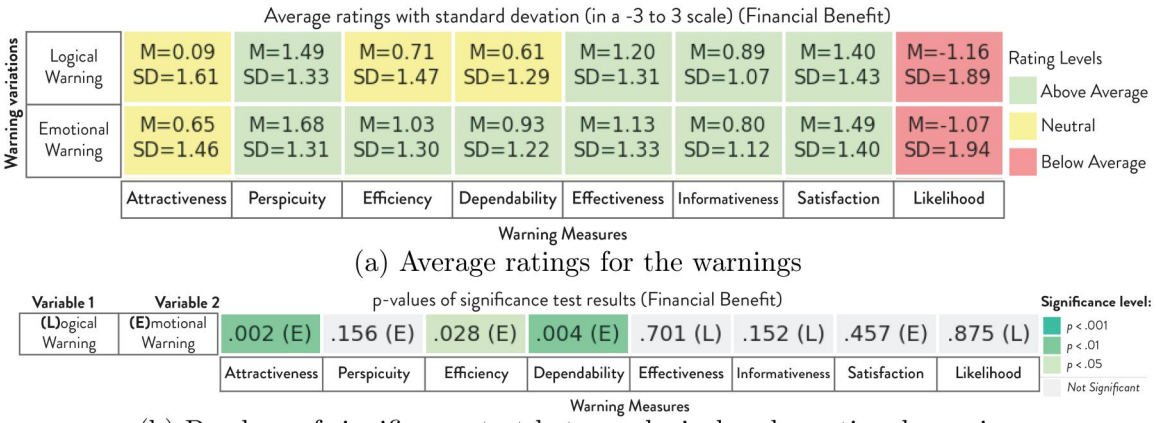


Figure 16. Average ratings for the clickbait post along with the likelihood to click with the warning (Financial benefit mental model).



(a) Average ratings for the warnings

(b) P-values of significance test between logical and emotional warning

Figure 17. Financial benefit mental model.

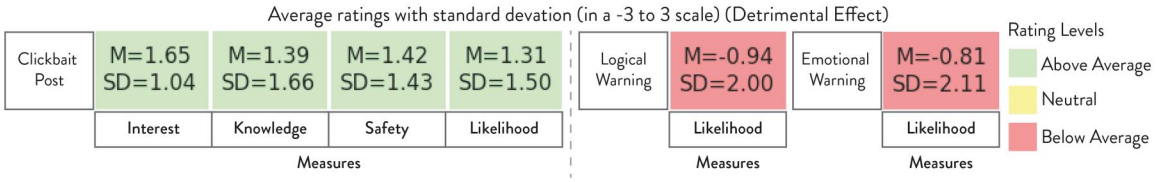


Figure 18. Average ratings for the clickbait post along with the likelihood to click with the warning (detrimental effect mental model).

with *Financial MM* found the emotional warning significantly more attractive, dependable, and efficient (Figure 17(b)). One participant commented,

> The graphics are really cute and descriptive and they actually tell you what can happen should you choose to click on the clickbait article.

4.2.3. Detrimental Effect (shortened as Detriment MM)

Detrimental Effect is the mental model where participants thought of clickbait as a tool to harm them by introducing malicious software to their devices. 8.86% of the participants had such a mental model. For instance, one participant commented,

> I think it's when you like the subject matter and want to explore, but it's actually bait that takes you to a virus or some kind of malware if you click on it.

These participants have high interest but comparatively lower likelihood to click on the post (Figure 18). That may be due to higher knowledge about the post and a lower perception of safety about the associated link. With the warnings, the likelihood scores decreased by more than two points (Figure 18). Both logical ($W = 113.5$, $p < .001$) and emotional ($W = 116.5$, $p < .001$) warnings significantly reduced the participant's likelihood to click on the post.

Participants with *Detriment MM* found the logical warning above average in perspicuity, effectiveness, and satisfaction (Figure 19(a)). Similarly, they rated the emotional one above average in perspicuity and satisfaction (Figure 19(a)). Since our warnings convey harm that these users already know about, they may have found the warnings to be less valuable, explaining the lower scores for many parameters. Moreover, there was no significant difference between the emotional and logical warnings in any of the parameters (Figure 19(b)).

4.3. Combined mental models (shortened as Combined MM)

We observed that 57.34% of participants defined clickbait based on its working and goals. For instance, one participant commented about both the working and goal of clickbait in explaining their understanding,

> To me, the term clickbait refers to an article or video with a misleading title and highly interesting title that entices people to click on it to read/see more. It is usually a controversial or popular topic, and the actual article or video is not about

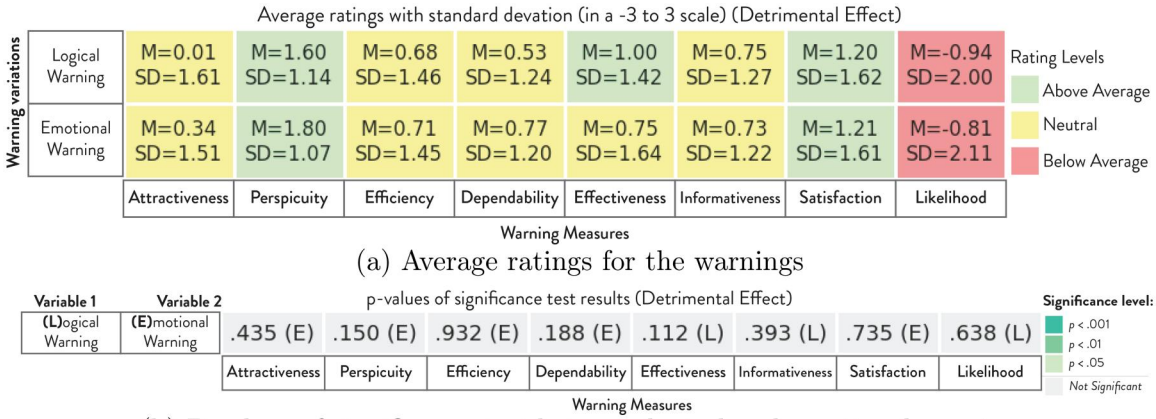


Figure 19. Detrimental effect mental model.

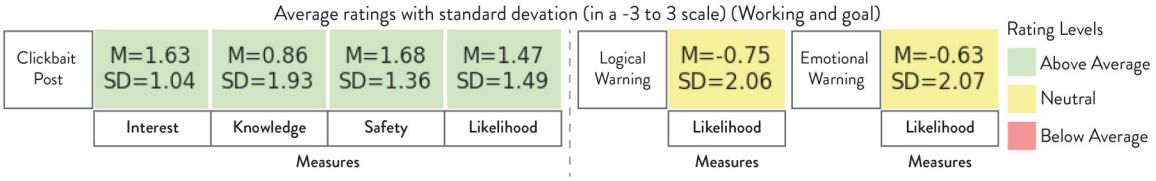


Figure 20. Average ratings for the clickbait post along with the likelihood to click with the warning (combined (working and goal) mental model).

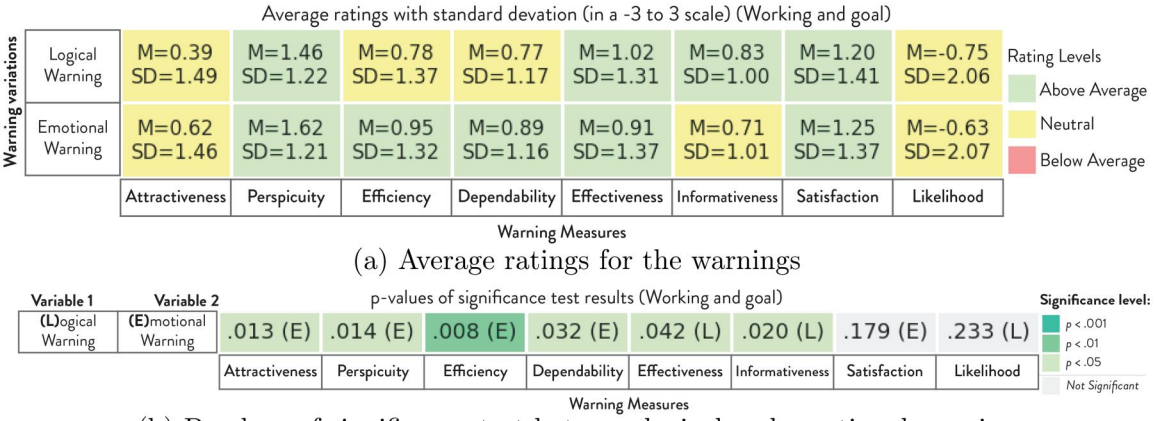


Figure 21. Combined (working and goal) mental model.

> whatever the title indicates. Basically, it's a misleading title that gets people to click because they're interested for some reason, but that's not what they find when they get to the link.

Participants with *Combined MM* have comparatively lower interest, lower likelihood to click, and higher knowledge about the post (Figure 20). Both logical and emotional warnings further decreased the likelihood of clicking on the post by more than two points (Figure 20). Significance tests revealed that both logical ($W = 3757.5$, $p < .001$) and emotional ($W = 4010.0$, $p < .001$) warnings reduced the participant's likelihood to click on the post.

These participants found the logical warning above average in perspicuity, effectiveness, informativeness, and satisfaction (Figure 21(a)). Similarly, the emotional warning was rated above average in perspicuity, efficiency, dependability, effectiveness, and satisfaction (Figure 21(a)). The participants found the emotional warning significantly more attractive, dependable, understandable, and efficient (Figure 21(b)). One participant said,

> I like the cartoon aspect of it; it catches my attention. I like that it tells me in a straightforward way that the post is clickbait.

Similarly, they found the logical warning significantly better regarding effectiveness and informativeness (Figure 21(b)). One participant mentioned,

> The images were nice and informative. And the text below it explained why.

For ease of viewing, we have also summarized the likelihood of each mental model group to click on clickbait with and without warning in Table 4.

Table 4. Summary of the likelihood to click on clickbait with and without warning.

Mental models	Without warning Mean	Std	Logical warning Mean	Std	Emotional warning Mean	Std
Working of clickbait	**1.53**	**1.58**	**−0.70**	**2.11**	**−0.61**	**2.11**
Information sensationalization	1.47	1.51	−0.72	2.07	−0.56	2.06
Deceptive presentation	1.55	1.51	−0.70	2.09	−0.60	2.11
Information camouflage	0.90	1.82	−1.55	1.88	−1.15	2.11
Goal of clickbait	**1.85**	**1.30**	**−0.36**	**2.24**	**−0.44**	**2.10**
Traffic increment	1.60	1.44	−0.63	2.10	−0.55	2.07
Financial benefit	1.25	1.49	−1.16	1.89	−1.07	1.94
Detrimental effect	1.31	1.50	−0.94	2.00	−0.81	2.11
Combined (working + goal)	**1.47**	**1.49**	**−0.75**	**2.06**	**−0.63**	**2.07**

Note: The bold text presents the sense-making lenses and the not bold ones present the mental models within these lenses.

5. Discussion

Our findings highlight the mental models of clickbait and its impact on users' perceptions of clickbait and countermeasures against it. In that section, we discuss the following: (1) using logic and emotion in warnings in §5.1, (2) the theoretical and practical implications of clickbait mental models on understanding and enhancing user knowledge and designing personalized interventions in §5.2, and (3) the necessity of education and awareness of clickbait in §5.3. Further, we highlight potential future works based on these implications throughout the section.

5.1. Leveraging logic and emotion in warnings

In our findings, both the logical and emotional warnings significantly reduced the likelihood of clicking on the post for users of all mental models compared to when there was no warning (see §4). Such a finding implies the importance of supporting users to reach informed decisions echoing prior works (Huang et al., 2015; Scott, 2021; Urakami et al., 2022) indicating the efficacy of beyond clickbait identification. While warnings from prior works informed users that a post is clickbait, they did not help them understand what clickbait is or why they should not click on it (Chakraborty et al., 2016). However, users can perceive clickbait as harmless as its consequences are not readily visible (Vance et al., 2017) indicating a need for informative warnings. In that regard, our findings show that logical and emotional warnings have their set of strengths and weaknesses across mental models. Our findings unveil the strength of emotional warnings in attractiveness, dependability, and efficiency across multiple mental models. Similarly, our study reveals the strength of logical warnings in effectiveness and informativeness across users of different mental models.

However, between the logical and emotional warnings, there was no significant difference in the likelihood of clicking on the post for any mental model. Such a lack of difference implies that even with their respective strengths and weaknesses regarding user experience, both warnings can effectively reduce the users' likelihood of clicking on the post. Further, our findings explain some of the potential reasons behind these strengths and weaknesses of the warnings. However, we urge future work to dive deeper to understand how user mental models and the concepts of logic and emotion influence the user experience in these parameters through qualitative methods such as interviews or focus group discussions (Baxter et al., 2015; Braun & Clarke, 2006).

5.2. Avenues for research using mental models

Our study provides the first look into the users' mental models of clickbait and contributes to understanding how these mental models impact their perceptions and behavior in the online setting. Here, our work aligns with the ideas of mental model and system image (Norman, 2013, 2014). In clickbait, attackers aim to obscure the system image, resulting in lack of clarification about its working and potential consequences through interventions. For instance, most consequences of clickbait are not visible, resulting in users not learning about them even after interacting with clickbait in the past. When the system image is obscured, Norman (2013) reports that mental models suffer. In our study, participants rarely had a comprehensive understanding of working of clickbait and its consequences to them (see §4). In fact, only 8.86% of participants knew about some of the consequences of clickbait in our study (see §4.2.3). Therefore, understanding mental models contributes to our understanding of the users' existing knowledge on clickbait and provides valuable insights in planning ways to enhance their mental models for instilling safer online security behavior.

In that regard, researchers and social media platforms can play an important role. While our study provides the first look into users' mental models of clickbait, we urge researchers to validate and add to the knowledge. In fact, several iterations of future works on how mental models can be enhanced (see §5.2.1) and how the knowledge about mental models can be leverage into creation of effective interventions are required (see §5.2.2). On the other hand, social media platforms need to adopt a more active approach on supporting users against clickbait. While these platforms are intent on stopping clickbait evidenced by their attempts at moderation (Gleicher, 2019; Roth & Harvey, 2018), they have rarely focused on supporting users. However, due to existing problems in moderation (D. Molina et al., 2021; Karande et al., 2021), users still encounter clickbait regularly. Therefore, in the short term, solutions such as crowdsourcing or using interventions similar to ones in our study that educate users about consequences of clickbait can be effective. Nevertheless, we urge social media platforms to take collaborative initiatives with clickbait

researchers to speed the production of viable mental model-based interventions in the long term.

5.2.1. Enhancing mental models through augmentation

Our findings show that users with combined mental models of working and goal of clickbait were less likely to click on the post than users with a singular mental model (working or goal only) (see §4.3, §4.1, and §4.2). Regarding the importance of these mental models, users who understand clickbait in multiple modes (working and goal) are better suited to deal with clickbait when there are no warnings. However, these same users rated the two warning variations poorly across most parameters for user experience and usability (see §4.3). Since these users understand clickbait using multiple modes, we believe they could have better knowledge about clickbait. Such knowledge may include the content of the warning, that is, the consequences of clicking on clickbait. As these users may already know about the warning content, we speculate that they could have found the warnings less helpful, resulting in poor ratings for user experience. Irrespective of the ratings, these users were more inclined to avoid clickbait without any warnings.

Such a finding highlights the importance of augmenting multiple mental models to create a more comprehensive model that increases users' capacity to identify and avoid clickbait (Kaptein et al., 2015; Liu et al., 2016). Warnings designed in our study to support users by increasing their understanding of clickbait can provide the initial direction for augmenting user mental models, resulting in safer online behavior (Kaptein et al., 2015; Liu et al., 2016). Based on these directions, we suggest researchers to focus on designs and interventions that help augment mental models in future (creating a complex and comprehensive mental model by introducing multiple simpler mental models). However, we acknowledge augmentation may be only one of many solutions to enhance mental models and urge researchers to explore new and innovative ideas in these directions.

5.2.2. Scopes of personalization

In designing mental model based interventions, our findings show that the different mental models can influence users' behaviors towards clickbait and the warnings against it (see Table 4 and §4). For instance, we can infer from our findings that users who understand clickbait based on goal alone are very likely to interact with clickbait compared to users of other mental models and may need stricter measures in the interventions. Even with warnings, these users are only neutral regarding clickbait interaction—these differences in clickbait interaction among user groups point towards a need for personalized warnings. Moreover, across the findings, users have different perceptions of the two warnings. For instance, we can infer that the users with information camouflage mental model prefer logical warnings (see §4.1.3). Similarly, some mental model groups (e.g., financial benefit, clickbait goal) have a slight preference for emotional warnings (see §4.2), and some (e.g., information sensationalization) have a mixed preference (see §4.1.1).

These findings support the narratives of prior works that highlight the importance of personalizing warnings based on users' abilities and knowledge (Kaptein et al., 2015; Liu et al., 2016). While our study focuses on logical and emotional warnings conveying harm, there are many unexplored techniques to change human behavior regarding clickbait (Abraham & Michie, 2008; Michie et al., 2013). We urge researchers to focus on these techniques to understand further scopes of personalization based on the mental models of the users. Further, we suggest future works to explore various methods to translate the mental models of users into effective interventions.

5.3. Need and plan for clickbait awareness and education

The users of the three mental models, information camouflage (see §4.1.3), financial benefit (see §4.2.2), and detrimental effects (see §4.2.3), had the lowest likelihood of clicking on the post without any warnings. These users also had the lowest likelihood of clicking on the post after seeing both the logical and the emotional warnings. However, these three groups had the lowest number of users, indicating that, like prior studies indicate (Geeng et al., 2020; Huang et al., 2015; Pennycook et al., 2022; Scott, 2021; Urakami et al., 2022), most users do not understand clickbait based on hiding information, getting financial benefit from them, or causing harm to them. These findings together point to a need for warnings against clickbait while instilling awareness and education among the users (Geeng et al., 2020; Huang et al., 2015; Pennycook et al., 2022; Urakami et al., 2022).

In achieving clickbait awareness and education, our findings suggest the importance of information about these three mental models since these users have the lowest likelihood of interacting with clickbait. We recommend future warnings and educational materials relating to clickbait to highlight the hiding of information, financial benefits for bad actors, and harm caused to users through clickbait. However, we urge researchers to validate the effectiveness of such warnings and materials before implementation. While our warnings focused on warnings conveying harm, there is still much scope to explore warnings and designs focused on the remaining two mental models, which our findings point to be effective against clickbait. Here, we urge the researchers to explore these directions and the social media platforms to implement solutions that focus on supporting and educating social media users. While these platforms have mostly focused on moderation, supporting and educating users can create a resilient community and can strengthen the targets of social engineering attacks, humans.

5.4. Limitations and future work

Our study presents mental models of clickbait based on the responses of 722 participants. However, mental models represent a person's understanding of a concept and can be

non-exhaustive. Therefore, the mental models presented in the study should not be considered as exhaustive. In fact, our goal is not to exhaustively identify the mental models of clickbait but to contribute to the knowledge about users' understanding of clickbait and its impact on their online security behavior. Further, we believe user mental models are to some extent shaped by their environment, implying that mental models of users with different culture, norms, and literacy can also be different. Recent HCI research (Al-Ameen et al., 2021; Al-Ameen & Kocabas, 2020; Shahid et al., 2022; Shrestha et al., 2023) supports such differences and puts importance on looking beyond the Western contexts, where factors such as cultural background, literacy rate, and economic condition could impact users' perceptions and behavior. Therefore, we suggest future studies involve participants from diverse geographic regions, including developing countries, to understand differences in users' mental models of clickbait and its impact on their perceptions and behavior towards clickbait and warnings against it.

In the online study, the warnings are presented in a mock platform where the interaction with the warning is the primary task. However, in social media, interaction with warnings is usually a secondary task. Therefore, future studies should consider the implementation and study of these warnings in a more realistic setting.

The non-significant results from the online study do not imply that such a relationship does not exist. Instead, it is only valid in the specified effect size and significance level. Further, quantitative results provide generalizable findings but can have gaps in making sense of some of the significant results. To understand the reasons behind these significant results, we suggest future works to conduct qualitative studies with users from different mental model groups identified in our study.

6. Conclusion

Overall, our findings contribute to the knowledge about users' mental models of clickbait (RQ1), where we unveil that users understand clickbait based on its working and goal. Under each of these mental models, users made sense of clickbait in diverse ways supporting the diverse nature of mental models. However, these mental models by themselves are rarely adequate to comprehend the dangers of clickbait and a need for interventions against clickbait are apparent. Our study evaluates two variations of such interventions that aims to not only classify a post as clickbait but also informs and educate users about clickbait and its consequences (RQ2). Our findings reveal the diverse needs of the user groups based on mental models and the scopes of personalization in these interventions. We conclude by highlighting the fact that mental models, in fact, have substantial influence on users' behavior towards clickbait and warnings against it and recommend personalization of interventions and augmentation of mental models to better prepare social media users against clickbait.

Notes

1. https://www.eccouncil.org/cybersecurity-exchange/ethical-hacking/understanding-preventing-social-engineering-attacks/
2. https://www.eccouncil.org/cybersecurity-exchange/ethical-hacking/understanding-preventing-social-engineering-attacks/
3. https://www.nngroup.com/articles/mental-models/
4. https://www.ueq-online.org/
5. https://www.mturk.com/worker/help

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Science Foundation under [Grant No. CNS-1949699].

ORCID

Ankit Shrestha http://orcid.org/0000-0002-9012-6146
Arezou Behfar http://orcid.org/0000-0001-8547-7868
Mahdi Nasrullah Al-Ameen http://orcid.org/0000-0002-5764-2253

References

Abraham, C., & Michie, S. (2008). A taxonomy of behavior change techniques used in interventions. *Health Psychology*, *27*(3), 379–387. https://doi.org/10.1037/0278-6133.27.3.379

Abu-Salma, R., & Livshits, B. (2020). Evaluating the end-user experience of private browsing mode [Paper presentation]. In Proceedings of the 2020 Chi Conference on Human Factors in Computing Systems (pp. 1–12), Honolulu, HI.

Agrawal, A. (2016). Clickbait detection using deep learning [Paper presentation]. 2016 2nd International Conference on Next Generation Computing Technologies (NGCT) (pp. 268–272), Dehradun, India.

Ajina, A. S., Javed, H. M. U., Ali, S., & Zamil, A. M. (2023). Fake or fact news? Investigating users' online fake news sharing behavior: The moderating role of social networking sites (SNS) dependency. *International Journal of Human–Computer Interaction*, 1–15. https://doi.org/10.1080/10447318.2023.2192108

Al-Ameen, M. N., & Kocabas, H. (2020). *"I cannot do anything": User's behavior and protection strategy upon losing, or identifying unauthorized access to online account* [Poster session]. Symposium on usable privacy and security.

Al-Ameen, M. N., Kocabas, H., Nandy, S., & Tamanna, T. (2021). "We, three brothers have always known everything of each other": A cross-cultural study of sharing digital devices and online accounts. *Proceedings on Privacy Enhancing Technologies*, *2021*(4), 203–224. https://doi.org/10.2478/popets-2021-0067

Aldawood, H., & Skinner, G. (2019). A taxonomy for social engineering attacks via personal devices. *International Journal of Computer Applications*, *178*(50), 19–26. https://doi.org/10.5120/ijca2019919411

Allen, J., Martel, C., & Rand, D. G. (2022). Birds of a feather don't fact-check each other: Partisanship and the evaluation of news in Twitter's Birdwatch crowdsourced fact-checking program [Paper presentation]. Chi Conference on Human Factors in Computing Systems (pp. 1–19), New Orleans, LA.

Amgoud, L., Bonnefon, J.-F., & Prade, H. (2007). The logical handling of threats, rewards, tips, and warnings. In *Symbolic and quantitative approaches to reasoning with uncertainty: 9th European Conference,*

ECSQARU 2007, Hammamet, Tunisia, October 31–November 2, 2007, Proceedings 9 (pp. 235–246). Springer.

Avery, J., Almeshekah, M., & Spafford, E. (2017). Offensive deception in computing. In International Conference on Cyber Warfare and Security (p. 23). Academic Conferences International Limited.

Babu, A., Liu, A., & Zhang, J. (2017). New updates to reduce clickbait headlines. Facebook Newsroom. https://about.fb.com/news/2017/05/news-feed-fyi-new-updates-to-reduce-clickbait-headlines/

Baxter, K., Courage, C., & Caine, K. (2015). *Understanding your users: A practical guide to user research methods* (2nd ed.). Morgan Kaufmann Publishers Inc.

Bhuiyan, M. M., Horning, M., Lee, S. W., & Mitra, T. (2021). NudgeCred: Supporting news credibility assessment on social media through nudges. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1–30. https://doi.org/10.1145/3479571

Bhuiyan, M. M., Zhang, K., Vick, K., Horning, M. A., & Mitra, T. (2018). FeedReflect: A tool for nudging users to assess news credibility on Twitter. *Companion of the 2018 ACM conference on computer supported cooperative work and social computing* (pp. 205–208). Association for Computing Machinery.

Bin Naeem, S., & Kamel Boulos, M. N. (2021). Covid-19 misinformation online and health literacy: A brief overview. *International Journal of Environmental Research and Public Health*, *18*(15), 8091. https://doi.org/10.3390/ijerph18158091

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Sage.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chakraborty, A., Paranjape, B., Kakarla, S., & Ganguly, N. (2016). Stop clickbait: Detecting and preventing clickbaits in online news media [Paper presentation]. 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 9–16). IEEE.

Chien, S.-Y., Yang, C.-J., & Yu, F. (2022). Xflag: Explainable fake news detection model on social media. *International Journal of Human–Computer Interaction*, *38*(18–20), 1808–1827. https://doi.org/10.1080/10447318.2022.2062113

Clark, B. (2013). *Relevance theory*. Cambridge University Press.

Cronkhite, G. L. (1964). Logic, emotion, and the paradigm of persuasion. *Quarterly Journal of Speech*, *50*(1), 13–18. https://doi.org/10.1080/00335636409382640

Dessart, L., & Standaert, W. (2023). Strategic storytelling in the age of sustainability. *Business Horizons*, *66*(3), 371–385. https://doi.org/10.1016/j.bushor.2023.01.005

D. Molina, M., Sundar, S. S., Rony, M. M. U., Hassan, N., Le, T., & Lee, D. (2021). Does clickbait actually attract more clicks? *Three* clickbait studies you must read [Paper presentation]. Chi Conference on Human Factors in Computing Systems (pp. 1, in Proceedings of the 2021–19), Yokohama, Japan.

Dumaru, P., Shrestha, A., Paudel, R., Haverkamp, C., McClain, M. B., & Al-Ameen, M. N. (2023). "… I have my dad, sister, brother, and mom's password": Unveiling users' mental models of security and privacy-preserving tools. *Information & Computer Security*, https://doi.org/10.1108/ICS-04-2023-0047

Ecker, U. K., Lewandowsky, S., & Tang, D. T. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, *38*(8), 1087–1100. https://doi.org/10.3758/MC.38.8.1087

Faris, R., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E., & Benkler, Y. (2017). *Partisanship, propaganda, and disinformation: Online media and the 2016 us presidential election* (p. 6). Berkman Klein Center Research Publication.

Fillenbaum, S. (1976). Inducements: On the phrasing and logic of conditional promises, threats, and warnings. *Psychological Research*, *38*(3), 231–250. https://doi.org/10.1007/BF00309774

Geeng, C., Yee, S., & Roesner, F. (2020). 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). Association for Computing Machinery.

Gleicher, N. (2019). Removing coordinated inauthentic behavior from china. *Facebook Newsroom*, *19*. https://about.fb.com/news/2019/08/removing-cib-china/

Hadnagy, C. (2010). *Social engineering: The art of human hacking*. John Wiley & Sons.

Hassan, N., Yousuf, M., Mahfuzul Haque, M., A. Suarez Rivas, J., & Khadimul Islam, M. (2019). Examining the roles of automation, crowds and professionals towards sustainable fact-checking [Paper presentation]. In Companion proceedings of the 2019 world wide Web Conference (pp. 1001–1006), San Francisco, CA.

Heuer, H., & Glassman, E. L. (2022). A comparative evaluation of interventions against misinformation: Augmenting the WHO checklist [Paper presentation]. Chi Conference on Human Factors in Computing Systems (pp. 1–21), New Orleans, LA.

Hu, D., & Apuke, O. D. (2023). Modeling the factors that stimulates the circulation of online misinformation in a contemporary digital age. *International Journal of Human–Computer Interaction*, 1–13. https://doi.org/10.1080/10447318.2023.2209839

Huang, Y. L., Starbird, K., Orand, M., Stanek, S. A., & Pedersen, H. T. (2015). Connected through crisis: Emotional proximity and the spread of misinformation online. Proceedings of the 18th *ACM* Conference on Computer Supported Cooperative Work & Social Computing (pp. 969–980). Association for Computing Machinery.

Huber, M., Kowalski, S., Nohlberg, M., & Tjoa, S. (2009). Towards automating social engineering using social networking sites [Paper presentation]. 2009 International Conference on Computational Science and Engineering (Vol. 3, pp. 117–124), Vancouver, BC, Canada.

Indrajit, R. E. (2017). Social engineering framework: Understanding the deception approach to human element of security. *International Journal of Computer Science Issues*, *14*(2), 8. https://doi.org/10.20943/01201702.816

Ipeirotis, P. G., Provost, F., & Wang, J. (2010). Quality management on amazon mechanical Turk [Paper presentation]. Proceedings of The ACM SIGKDD Workshop on Human Computation (pp. 64–67), Washington, DC.

Javed, R. T., Shuja, M. E., Usama, M., Qadir, J., Iqbal, W., Tyson, G., Castro, I., & Garimella, K. (2020). A first look at covid-19 messages on Whatsapp in Pakistan [Paper presentation]. 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 118–125), The Hague, Netherlands.

Johnston-Laird, P. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness* (No. 153.4 JOHm). Harvard University Press.

Kaiser, B., Wei, J., Lucherini, E., Lee, K., Matias, J. N., Mayer, J. (2021). Adapting security warnings to counter online disinformation. In 30th Usenix Security Symposium (Usenix Security 21)(pp. 1163–1180). USENIX Association.

Kang, R., Dabbish, L., Fruchter, N., & Kiesler, S. (2015). "My data just goes everywhere:" User mental models of the internet and implications for privacy and security. In Eleventh Symposium on Usable Privacy and Security (Soups)(pp. 39–52). USENIX Association.

Kaptein, M., Markopoulos, P., De Ruyter, B., & Aarts, E. (2015). Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles. *International Journal of Human-Computer Studies*, *77*, 38–51. https://doi.org/10.1016/j.ijhcs.2015.01.004

Karande, H., Walambe, R., Benjamin, V., Kotecha, K., & Raghu, T. (2021). Stance detection with Bert embeddings for credibility analysis of information on social media. *PeerJ Computer Science*, *7*, e467. https://doi.org/10.7717/peerj-cs.467

Kee, J., & Deterding, B. (2008). Social engineering: Manipulating the source. *GCIA Gold Certification.* https://www.giac.org/paper/gcia/2968/social-engineering-manipulating-source/115738

Khiralla, F. A. M. (2020). Statistics of cybercrime from 2016 to the first half of 2020. *International Journal of Computer Science Network*, *9*(5), 252–261.

Konstantinou, L., Panos, D., & Karapanos, E. (2024). Exploring the design of technology-mediated nudges for online misinformation.

International Journal of Human–Computer Interaction, 1–28. https://doi.org/10.1080/10447318.2023.2301265

Krombholz, K., Hobel, H., Huber, M., & Weippl, E. (2015). Advanced social engineering attacks. *Journal of Information Security and Applications*, *22*, 113–122. https://doi.org/10.1016/j.jisa.2014.09.005

Kumaraguru, P., Sheng, S., Acquisti, A., Cranor, L. F., & Hong, J. (2010). Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology*, *10*(2), 1–31. https://doi.org/10.1145/1754393.1754396

Kung, F. Y., Kwok, N., & Brown, D. J. (2018). Are attention check questions a threat to scale validity? *Applied Psychology*, *67*(2), 264–283. https://doi.org/10.1111/apps.12108

Lan, X., Wu, Y., Shi, Y., Chen, Q., & Cao, N. (2022). Negative emotions, positive outcomes? exploring the communication of negativity in serious data stories [Paper presentation]. Chi Conference on Human Factors in Computing Systems (pp. 1–14), New Orleans, LA.

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Li, X., Zhou, J., Xiang, H., & Cao, J. (2022). Attention grabbing through forward reference: An ERP study on clickbait and top news stories. *International Journal of Human–Computer Interaction*, 1–16. https://doi.org/10.1080/10447318.2022.2158262

Liu, B., Andersen, M.S., Schaub, F., Almuhimedi, H., Zhang, S.A., Sadeh, N., Agarwal, Y., & Acquisti, A. (2016). Follow my recommendations: A personalized privacy assistant for mobile app permissions. *In* Soups 2016-Proceedings of the 12th Symposium on Usable Privacy and Security(pp. 27–41). USENIX Association.

Marwick, A. E., & Lewis, R. (2017). *Media manipulation and disinformation online*. Data and Society.

Mayer, R. E. (2014). Introduction to multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 1–24). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.002

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, J., & Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine*, *46*(1), 81–95. https://doi.org/10.1007/s12160-013-9486-6

Moreno, R., & Valdez, A. (2005). Cognitive load and learning effects of having students organize pictures and words in multimedia environments: The role of student interactivity and feedback. *Educational Technology Research and Development*, *53*(3), 35–45. https://doi.org/10.1007/BF02504796

Murnane, E. L., Jiang, X., Kong, A., Park, M., Shi, W., Soohoo, C., Vink, L., Xia, I., Yu, X., Yang-Sammataro, J., & Young, G. (2020). Designing ambient narrative-based interfaces to reflect and motivate physical activity [Paper presentation]. Proceedings of the 2020 Chi Conference on Human Factors in Computing Systems (pp. 1–14), Honolulu, HI..

Norman, D. (2013). *The design of everyday things* (revised and expanded edition). Basic books.

Norman, D. A. (2014). Some observations on mental models. In *Mental models* (pp. 15–22). Psychology Press.

Oates, M., Ahmadullah, Y., Marsh, A., Swoopes, C., Zhang, S., Balebako, R., & Cranor, L. F. (2018). Turtles, locks, and bathrooms: Understanding mental models of privacy through illustration. *Proceedings on Privacy Enhancing Technologies*, *2018*(4), 5–32. https://doi.org/10.1515/popets-2018-0029

O'Donnell, A. (2018, May). What is clickbait?: What's really happening when you click that link to finish an irresistible story. https://www.lifewire.com/the-dark-side-of-clickbait-2487506.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Paivio, A. (2006). *Mind and its evolution: A dual coding theoretical interpretation*. Lawrence Erlbaum Associates, Inc.

Paudel, R., Shrestha, A., Dumaru, P., & Al-Ameen, M. N. (2023). "It doesn't just feel like something a lawyer slapped together." Mental-model-based privacy policy for third-party applications on Facebook. *Companion publication of the 2023 conference on computer supported cooperative work and social computing* (pp. 298–306). Association for Computing Machinery.

Peck, A. (2020). A problem of amplification: Folklore and fake news in the age of social media. *Journal of American Folklore*, *133*(529), 329–351. https://doi.org/10.5406/jamerfolk.133.529.0329

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on amazon mechanical Turk. *Behavior Research Methods*, *46*(4), 1023–1031. https://doi.org/10.3758/s13428-013-0434-y

Pennycook, G., McPhetres, J., Bago, B., & Rand, D. G. (2022). Beliefs about COVID-19 in Canada, the United Kingdom, and the United States: A novel test of political polarization and motivated reasoning. *Personality & Social Psychology Bulletin*, *48*(5), 750–765. https://doi.org/10.1177/01461672211023652

Pine, K. H., Lee, M., Whitman, S. A., Chen, Y., & Henne, K. (2021). Making sense of risk information amidst uncertainty: Individuals' perceived risks associated with the covid-19 pandemic [Paper presentation]. In Proceedings of the 2021 Chi Conference on Human Factors in Computing Systems (pp. 1–15), Yokohama, Japan.

Redmiles, E. M., Chachra, N., & Waismeyer, B. (2018). Examining the demand for spam: Who clicks? In Proceedings of the 2018 Chi Conference on Human Factors in Computing Systems (p. 212), Montreal, QC.

Rides, K. (2017, August). Clickbait malware sites. https://www.linkedin.com/pulse/clickbait-malware-sites-kris-rides/.

Roth, Y., & Harvey, D. (2018, June). *How twitter is fighting spam and malicious automation*. Twitter [blog]

Safety, T. (2019, August 19). *Information operations directed at Hong Kong. Twitter Blog.*

Schul, Y. (1993). When warning succeeds: The effect of warning on success in ignoring invalid information. *Journal of Experimental Social Psychology*, *29*(1), 42–62. https://doi.org/10.1006/jesp.1993.1003

Scott, K. (2021). You won't believe what's in this paper! Clickbait, relevance and the curiosity gap. *Journal of Pragmatics*, *175*, 53–66. https://doi.org/10.1016/j.pragma.2020.12.023

Shahid, F., Kamath, S., Sidotam, A., Jiang, V., Batino, A., & Vashistha, A. (2022). "It matches my worldview": Examining perceptions and attitudes around fake videos. In *Chi conference on human factors in computing systems* (pp. 1–15), Association for Computing Machinery.

Sharevski, F., Treebridge, P., Jachim, P., Li, A., Babin, A., & Westbrook, J. (2022). Socially engineering a polarizing discourse on Facebook through malware-induced misperception. *International Journal of Human–Computer Interaction*, *38*(17), 1621–1637. https://doi.org/10.1080/10447318.2021.2009671

Shrestha, A., Paudel, R., Dumaru, P., & Al-Ameen, M. N. (2023). Towards improving the efficacy of windows security notifier for apps from unknown publishers: The role of rhetoric. International Conference on Human-Computer Interaction (pp. 101–121). Springer Nature Switzerland.

Shrestha, A., Sharma, T., Saha, P., Ahmed, S. I., & Al-Ameen, M. N. (2023). A first look into software security practices in Bangladesh. *ACM Journal on Computing and Sustainable Societies*, *1*(1), 1–24. https://doi.org/10.1145/3616383

Simmons, A. (2019). *The story factor: Inspiration, influence, and persuasion through the art of storytelling*. Basic books.

Souza, F. (2015, June). Analyzing a Facebook clickbait worm. https://blog.sucuri.net/2015/06/analyzing-a-facebook-clickbait-worm.html.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, *57*(1), 31–95. https://doi.org/10.1016/0010-0277(95)00666-m

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (vol. 142). Citeseer.

Sylvia Chou, W.-Y., Gaysynsky, A., & Cappella, J. N. (2020). *Where we go from here: Health misinformation on social media* (vol. 110, No. S3). American Public Health Association.

Tasnim, S., Hossain, M. M., & Mazumder, H. (2020). Impact of rumors and misinformation on covid-19 in social media. *Yebang Uihakhoe Chi* [Journal of Preventive Medicine and Public Health], *53*(3), 171–174. https://doi.org/10.3961/jpmph.20.094

Thatcher, A., & Greyling, M. (1998). Mental models of the internet. *International Journal of Industrial Ergonomics*, *22*(4–5), 299–305. https://doi.org/10.1016/S0169-8141(97)00081-4

Urakami, J., Kim, Y., Oura, H., & Seaborn, K. (2022). Finding strategies against misinformation in social media: A qualitative study [Paper presentation]. Chi Conference on Human Factors in Computing Systems Extended Abstracts (pp. 1–7), New Orleans, LA.

Vance, A., Kirwan, B., Bjornn, D., Jenkins, J., Anderson, B. B. (2017). What do we really know about how habituation to warnings occurs over time? A longitudinal FMRI study of habituation and polymorphic warnings. Proceedings of the 2017 Chi Conference on Human Factors in Computing Systems (pp. 2215–2227). Association for Computing Machinery.

Vasudeva, F., & Barkdull, N. (2020). Whatsapp in India? A case study of social media related lynchings. *Social Identities*, *26*(5), 574–589. https://doi.org/10.1080/13504630.2020.1782730

Wineburg, S., & McGrew, S. (2019). Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information. *Teachers College Record: The Voice of Scholarship in Education*, *121*(11), 1–40. https://doi.org/10.1177/016146811912101102

Wittes, B., Poplin, C., Jurecic, Q., & Spera, C. (2016). *Sextortion: Cybersecurity, teenagers, and remote sexual assault* (pp. 1–47). Center for Technology Innovation at Brookings.

Woolbert, C. H. (1918). The place of logic in a system of persuasion. *Quarterly Journal of Speech*, *4*(1), 19–39. https://doi.org/10.1080/00335631809360643

Wu, J., & Zappala, D. (2018). When is a tree really a truck? Exploring mental models of encryption. In Fourteenth Symposium on Usable Privacy and Security (Soups)(pp. 395–409). USENIX Association.

Xiang, H., Zhou, J., & Wang, Z. (2023). Reducing younger and older adults' engagement with covid-19 misinformation: The effects of accuracy nudge and exogenous cues. *International Journal of Human–Computer Interaction*, 1–16. https://doi.org/10.1080/10447318.2022.2158263

Young, I. (2008). *Mental models: Aligning design strategy with human behavior*. Rosenfeld Media.

Zeng, E., Kohno, T., & Roesner, F. (2020). Bad news: Clickbait and deceptive ads on news and misinformation websites. In *Workshop on technology and consumer protection* (pp. 1–11).

Zhang, Y., Suhaimi, N., Yongsatianchot, N., Gaggiano, J. D., Kim, M., Patel, S. A., Sun, Y., Marsella, S., Griffin, J., & Parker, A. G. (2022). Shifting trust: Examining how trust and distrust emerge, transform, and collapse in COVID-19 information seeking [Paper presentation]. Chi Conference on Human Factors in Computing Systems (pp. 1–21), New Orleans, LA.

Zheng, H.-T., Chen, J.-Y., Yao, X., Sangaiah, A. K., Jiang, Y., & Zhao, C.-Z. (2018). Clickbait convolutional neural network. *Symmetry*, *10*(5), 138. https://doi.org/10.3390/sym10050138

Zhou, Y. (2017). Clickbait detection in tweets using self-attentive network. *CoRR*. http://arxiv.org/abs/1710.05364.

About the authors

Ankit Shrestha is a PhD candidate and research assistant at the PIXEL (Privacy, desIgn, and user eXperiencE Lab) in Computer Science department of Utah State University. His research interests lie in the boundary of human computer interaction and privacy including a focus on behavior changing interventions.

Arezou Behfar, a Computer Science PhD candidate at Utah State University, specializes in user experience research and usability testing. She engages in collaborative HCI problem-solving at the PIXEL Lab, utilizing methods like design, prototyping, and interviews.

Mahdi Nasrullah Al-Ameen leads the PIXEL in the Computer Science department of Utah State University. He completed his PhD from University of Texas, Arlington in 2016. His research work focuses on systemizing the human and societal factors that impact people's secure and privacy-preserving use of a technology.