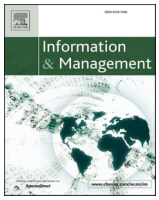
Information & Management 63 (2026) 104256

Contents lists available at ScienceDirect

Information & Management

journal homepage: www.elsevier.com/locate/im



The information processing of fake news: How intervention order influences perception over time

Christian Olenberger [a], Manfred Schoch [a,b,c,*], Lena Utz [c]

[a] FIM Research Center, Augsburg, Germany
[b] Esslingen University of Applied Sciences, Esslingen am Neckar, Germany
[c] Branch Business & Information Systems Engineering of the Fraunhofer FIT, Augsburg, Germany

ARTICLE INFO

Keywords:
Fake news
Information processing
Mental model
Social media
Intervention

ABSTRACT

Our ability to recognize fake news is flawed, with serious negative consequences for individuals, organizations, and society. Researchers and practitioners have developed different interventions against fake news, such as warnings and additional information, with mixed results. Yet social media users remain vulnerable to fake news because they consume information hedonically. To design more effective interventions, we must better understand how social media users process information cognitively. Drawing on dual process theory, we hypothesize that the sequence of exposure to information matters, and that it is beneficial to present interventions with warnings and additional information after exposure to fake news rather than before. We argue that questioning false information typically requires more cognitive resources than users are willing to allocate during casual browsing. In a two-week multiphase experiment, participants were exposed to fake news and interventions in different sequences. We compared the results across groups and multiple time points through statistical testing. The results show that introducing an intervention after a fake news story leads to a lower adoption of fake news into social media users' mental models and that the fake news's believability is then lowest. However, the effect fades over time without repeated exposure to interventions. Our findings offer new insights into how timing and cognitive effort influence the effectiveness of fake news interventions. For practitioners, we provide recommendations for designing effective interventions against fake news.

1. Introduction

Knowledge is power. The phrase attributed to Francis Bacon in 1597 remains more relevant than ever. People continue to seek knowledge, often by consuming news from a variety of sources. In recent years, however, the ways in which news is produced and consumed have changed fundamentally. Most importantly, social media platforms have become essential to both the dissemination and consumption of information [67]. However, in the absence of robust quality control mechanisms, these platforms remain susceptible to the spread of disinformation and fake news. For example, following the recent restructuring of X (formerly Twitter), the EU warned that the social network's existing measures against fake news and hate speech are inadequate to meet the requirements of EU law [46].

Under the Digital Service Act, the EU requires social media providers to remove illegal content or face the possibility of severe financial penalties [71]. However, unlike other harmful content types, fake news is not always clearly illegal. As a result, social media platform providers are often faced with the difficult task of determining whether to remove such content or to regard it as protected under the principle of freedom of expression. Moreover, the scope of the Digital Service Act is limited to the EU, which complicates enforcement across global social networks.

The increasingly negative influences of fake news are evident in public discourse on social and political events [21,55]. Fake news refers to demonstrably false information that is explicitly used to deceive consumers [54,68]. Its spread may lead individuals to accept inaccurate information as true, which has been shown to influence opinions and actions [3,25]. There are numerous prominent examples of the real-world consequences of disinformation campaigns [52,63]. The spread of fake news is expected to continue to increase, and its influence is therefore expected to grow – increasingly affecting individuals, organizations, and society.

A key factor contributing to the rise of fake news is the popularity and widespread use of social media platforms [28]. In traditional media

* Corresponding author.
E-mail address: manfred.schoch@fim-rc.de (M. Schoch).

https://doi.org/10.1016/j.im.2025.104256
Received 22 April 2024; Received in revised form 23 September 2025; Accepted 24 September 2025
Available online 25 September 2025

environments, journalists function as gatekeepers and are tasked with upholding the standards for the quality of information published in official media outlets. In contrast, social media enables users to create and disseminate information without adhering to journalistic principles or editorial oversight. As a result, large volumes of unverified and potentially misleading information circulate on social media platforms [65].

The trend is particularly problematic because the human ability to detect fake news is flawed. Individuals tend to accept fake news as true without critical scrutiny [38]. Moreover, confirmation bias leads users to perceive content as more credible when it aligns with their preexisting beliefs, opinions, and mental models [19,25,32]. Additional factors – such as previous exposure [59], rhetorical devices [9,45], familiarity [18], cognitive laziness [60], affective engagement [42], or mood [18] – have further negative impacts on our ability to evaluate fake news. Owing to these vulnerabilities, social media platforms have a responsibility to assist users in navigating the high volume of fake news.

Social media platform operators and researchers have recognized the growing threat of this problem, promoting the development and evaluation of different intervention types. These interventions are designed to support users in verifying information and to provide additional information surrounding the fake news stories. Although initial positive effects of such interventions have been observed, their overall effectiveness remains unclear [55]. For instance, previous research has shown that users frequently disregard warnings of fake news and source credibility ratings due to confirmation bias [19,34,49]. Further, Burel et al. [13], Wang et al. [74], and Pal and Banerjee [57] suggest that corrective information (such as fact-checks and the rebuttal of rumors) to reduce the spread of disinformation has limited effectiveness. Yet, other studies indicate that warning labels are effective in fighting fake news. Such labels can increase user skepticism during exposure to fake news [38]. Also, the inclusion of additional verified information can draw attention to the deceptive elements of a fake news story [11]. To better understand what makes these interventions effective against fake news, further research into how social media users process information is necessary.

Individuals use personal mental models to represent complex phenomena and events internally [38]. Research suggests that novelty and timing of information exposure are key factors influencing the formation of mental models. For instance, users tend to adopt new information – whether fake or accurate – into their mental models [11,38]. While browsing through social media, users tend to minimize cognitive effort and rarely engage in critical evaluation of the content they see. This is because social media are often used primarily as entertainment media that confront users with high volumes of information. Such conditions are problematic, as correcting previously accepted information as correct requires substantial cognitive effort [38]. In sum, social media users tend to adopt information as true easily and subsequently struggle to revise their existing mental models. This highlights the importance of the sequences in which individuals encounter information and its influence on mental model formation.

Theoretical arguments on the timing of interventions generally fall into two camps. The first is grounded in inoculation theory [39,40]. Inoculation theory posits that information presented before exposure to fake news can serve as a cognitive "vaccine," making users more resistant to disinformation. Interventions – such as warnings or additional information – are most effective when presented before exposure to a fake news story. This is because the first piece of information encountered serves as an anchor that influences subsequent judgments [33]. This approach aims to foster skepticism toward false information [39].

In contrast, the second camp draws on research into the misinformation effect [15], which suggests that presenting fake news first, followed by an intervention, may be more effective. According to this view, newer information is more readily retrieved and can overwrite earlier memories [15].

To ensure the effective use of interventions against fake news, it is essential to understand the underlying information processing behavior and its possible consequences better – particularly over time and in terms of its long-term effectiveness. Yet, to date, such empirical research contributions are missing. This is congruent with a recent call for longitudinal empirical IS research into interventions against fake news [24]. Thus, we ask

RQ1. What influence does the order of information presented to social media users as part of an intervention against fake news have on their immediate perceptions of fake news and how do these perceptions evolve over time?

RQ2. What is the impacts of the order of information on the integration of fake news and verified information into users' mental models?

To answer these questions, we combined two intervention strategies (warnings and verified information) and investigated whether the order in which they were presented influenced information processing and mental model formation. Building on existing research, we derived theoretical hypotheses and tested them through a longitudinal online study. Our results reveal that the order of interventions significantly influences the perception of fake news. Specifically, presenting a warning and verified information after exposure to fake news leads users to assess the false information as less believable and to adopt fake news less into their mental models. Further, when verified information is presented after the exposure to fake news, the verified information tends to persist longer in social media users' mental models. However, our results indicate that the effectiveness shrinks over time as users tend to forget the information that was displayed to them.

Our findings contribute to both research and practice. For researchers, we provide new insights on the underlying information processing processes during interventions against fake news. In contrast to previous studies (e.g., [58]), which have typically focused on immediate effects or single intervention types, our work adopts a longitudinal perspective to investigate how different sequences of interventions influence the persistence of information over a two-week period. For practitioners, our results provide evidence-based recommendations for designing more effective interventions against fake news. By demonstrating the benefits of combining warning messages with verified information and highlighting the importance of intervention orders' long-term effects, we present actionable strategies than can help mitigate the spread and negative consequences of fake news for individuals, organizations, and society.

2. Theoretical background

2.1. Fake news and how users create mental models

Two key characteristics define fake news: inauthenticity and deception [23]. Fake news contains verifiably false information that is deliberately spread [68]. Disinformation is essentially the same as fake news, as both involve the deliberate spread of false information in order to deceive [2]. Unlike misinformation, which is incorrect information shared without an intention to deceive, fake news is crafted to mislead [2]. Rumors are potentially unverified pieces of information that spread informally and may not have an explicit intention to deceive [57]. Unlike a lie, which typically involves a single statement or claim, fake news encompasses entire stories, articles, or reports designed to appear legitimate [68]. Over time, fake news can lead to enduring myths and false beliefs becoming embedded in culture [66].

Fake news can take various forms, including texts, images, audio

files, and videos that have been altered or taken out of context [26,54,68]. A sensation-oriented design often attracts potential readers' attention [65]. A significant problem with fake news is that users believe it without recognizing that it is not based on facts. For instance, we tend to accept information when no contradictory information is available [38]. Even when contradictory information is presented, as Pal et al. [57] demonstrated in their study on online health rumors, users are more likely to believe a rumor than a counter-rumor.

To understand the processes through which fake news becomes ingrained in a user's memory, mental models must be understood. "A mental model is an internal representation or simulation that people carry in their minds of how the world works" ([35], p. 16). Mental models are cognitive constructs that are built dynamically and are based on a person's memory and the current context in which they operate [29]. Mental models go beyond merely retaining isolated facts—they represent interconnected, dynamic cognitive structures that enable individuals to interpret information, draw conclusions, and make predictions [29,35]. While information retention refers to the ability to recall specific details, mental model formation involves actively processing and organizing this information into a coherent framework [38]. This distinction is crucial, as an effective intervention against fake news should not only enhance information retention but also facilitate the restructuring of mental models.

However, owing to the constraints of working memory and variability between personal experiences, these models are often functional but may lack precision or completeness [35]. Further, mental models are influenced by the information that individuals can access at the time of use, which can further impact their accuracy and completeness [35]. Thus, contact with fake news often leads to false information becoming part of a person's current and future mental model. Questioning or correcting mental models involves substantial cognitive effort [38]. Recent research shows that interventions that encourage deeper cognitive engagement – beyond just clicking – help users reconsider false beliefs [70].

Dual process theory distinguishes between two cognitive systems used for information assessment. System 1 processes are fast, automatic, effortless, emotionally charged, and heuristic. In contrast, System 2 processes are slower, reflective, controlled, efficient, and rational [30]. However, because our cognitive abilities are limited [36], we primarily use System 1 thinking in our daily lives [72]. This is especially true for social media, as the high information density in social media (many posts with different information in brief intervals) would otherwise result in high use of cognitive resources [44]. This explains our tendency not to evaluate fake news in social media properly, but to accept it directly into our mental models [38].

Reliance on System 1 thinking may also explain why users are more likely to hold on to a rumor, even when they are exposed to both a rumor and a counter-rumor [33,57]. The fast, automatic processing of information often makes us easy to influence, owing to rhetoric's persuasive and manipulative elements [53], such as the original rumor. Questioning false information usually uses more cognitive resources than we are willing to allocate.

2.2. Interventions against fake news

The problem of fake news has already been recognized in research and practice, and different intervention measures have been presented. Ng et al. [55] identified three classes of measures. First, community-level measures motivate users to identify fake news and stop its spread [55]. For instance, Gimpel et al. [23] showed that social media platforms can use social norms to encourage users to report fake news. Second, account-level measures stop the spread of fake news by limiting possible actions by malicious user accounts [55]. This includes blocking accounts [37] or disabling the function of sharing content [21]. Third, content-level measures highlight information or provide additional corrective information about fake news [55]. There has been research into different warning label types (e.g., [50,51]), ratings of sources (e.g., [32]), fact-checking [13], the rebuttal of rumors [13], and additional information (e.g., [22]).

In particular, flagging with warning labels has been studied widely [21]. This approach seeks to nudge people to engage more deeply with suspect content rather than relying on heuristics. For instance, Garrett and Poulsen [20], as well as Moravec et al. [49], showed that flagging can reduce belief in and intention to share fake news. However, the overall research results have been mixed. While Arendt et al. [5] showed that warning labels reduce the perceived accuracy of fake news among politically left-leaning individuals, they could not demonstrate this effect for politically right-leaning individuals. Ross et al. [64] and Moravec et al. [51] also failed to show a significant effect of flagging with warning labels on perceptions of fake news. While they demonstrated that warning labels can increase cognitive activity, they could not observe any effect on perceived believability owing to confirmation bias [51]. Further, warning labels can backfire. For instance, research shows that exposure to warning labels can increase general skepticism, making even truthful news more likely to be viewed as manipulated [64].

Another approach is to provide additional verified information and related articles to help people assess the accuracy and correct possible misconceptions. Qualified institutions, such as fact-checking organizations, can manually provide verified information [43]. Another complementary approach focuses on behavioral strategies: Prompting users to form implementation intentions, specific "if–then" plans, has been shown to increase fact-checking before sharing content significantly [6]. The use of artificial intelligence to identify fake news [17,62] and subsequent extraction of verified information from related articles from credible sources is also an option [79]. Gimpel et al. [22] showed that the additional provision of related articles improves users' identification of fake news in social media.

In contrast, however, studies have also shown that corrections can have a negative effect and can make us more likely to believe fake news, as preference-incongruent information can lead us to develop arguments in favor of our own opinions, strengthening our beliefs [56].

In sum, research has provided no clear evidence of the effectiveness of interventions against fake news. While the research has contributed greatly to our understanding of different interventions' effectiveness, including a recent systematic review by Hartwig et al. [27], we need to deepen our understanding of the cognitive processes that influence user perceptions.

2.3. Combining the flagging of fake news with additional verified information

We use mental models to process complex situations and events. Mindlessly scrolling (i.e., processing in System 1 instead of reflecting critically on the provided information) leads to the phenomenon that false information can be falsely and subconsciously adopted and accepted into mental models [11,38]. While fake news warning labels can increase attention in the short term [48], social media users cannot form a correct mental model of a situation by receiving only a warning label and no other information. Thus, we are forced to form a mental model based on false information and to tag it with a negation tag [38]. When we later access our mental model on the issue, we can only retrieve the false information and the negation tag. However, the memory of the negation tag can be lost over time because it is less frequently reinforced and is processed as secondary information, compared with the original false information. The initial false information is encoded more strongly owing to its higher salience, while the

negation tag is more vulnerable to cognitive biases and less robust reinforcement, making it easier to forget [24,38]. Thus, only the false information remains in the mental model, and it may be perceived as the truth. For interventions to be successful and sustainable, it is necessary to enable individuals to build a correct mental model.

We further drew on the results of research into propaganda. In our context, the psychological inoculation process describes two key elements: First, a warning label increases people's attention and caution and puts them in a state of resistance [40]; second, additional verified information allows them to compare false information with verified information [39]. It is crucial to combine these two elements, because a warning label increases caution and motivates us to process information actively, that is, to use System 2 thinking to create a correct mental model based on verified information.

3. Hypothesis development

3.1. The influence of information sequence on information processing

In addition to the informational content itself, the order in which information is presented plays a crucial role in information processing. From a cognitive consistency perspective, new information that contradicts existing knowledge leads to inconsistencies in processing information [38]. Resolving these inconsistencies requires comparing the prior information with the new input and deriving new meaning – a laborious process that consumes cognitive resources and requires user motivation [38]. Consequently, System 1 thinking often leads individuals to dismiss the new, contradictory information in favor of maintaining cognitive consistency. This tendency to reject conflicting information reinforces and perpetuates confirmation bias.

In contrast, research on misinformation suggests that individuals often adopt incorrect new information over the accurate details they initially encountered: "participants tended to report incorrect new information rather than the correct information they originally witnessed" ([15], p. 585). This phenomenon highlights that new information is cognitively more readily available and can overwrite old information [7,12]. Multiple studies [40,78] have demonstrated this effect, emphasizing how information processing is susceptible to recent misinformation.

Previous research suggests that when we encounter information that does not fit existing information, a state of cognitive dissonance is created [41]. "Cognitive dissonance occurs when users are presented with two pieces of conflicting information that both cannot be true" ([51], p. 10). Cognitive dissonance creates a conflict, which requires cognitive effort to resolve [35,38]. Since we use social media primarily for entertainment and pleasure [48], and therefore mainly use System 1 thinking, we try to avoid cognitive effort [41]. We apply a mental shortcut [80]. Thus, if users first see warning labels and verified information and then fake news, they engage less with the differences between the fake news and the verified information. Since we receive the correct information first, there is no motivation to compare the false content and to resolve contradictions. The willingness to use System 2 thinking is reduced in order to save cognitive resources, and the wrong information can become an undifferentiated part of our mental model [47,80].

An alternative would be to first display fake news, followed by warning labels and verified information. In this sequence, individuals first process false information and then see warnings and alternative information. This order is theorized to induce cognitive dissonance, prompting users to evaluate and engage more deeply with the conflicting information; thus, they will engage in analytical System 2 thinking [47]. Zúñiga et al. [80] emphasize that interventions enhancing cognitive engagement can help mitigate the spread of fake news. Similarly, Mirhoseini et al. [47] showed that encouraging users to process information more analytically reduces their susceptibility to fake news. Alon et al. [4] demonstrated that critical thinking lowers the perceived accuracy of fake news and influences subsequent online behavioral intentions positively. Analytical thinking is therefore necessary to prevent the integration of false information into mental models. In contrast, when verified information is presented first, users may lack motivation to scrutinize subsequent false content and have no incentive to resolve the potential conflict. Further, a negation tag can be lost over time [24,38], diminishing users' ability to distinguish between verified and false information.

However, the cognitive dissonance from receiving fake news first nudges us to invest cognitive effort [51]. This process facilitates the identification of fake news and encourages prioritization of verified information, thereby decreasing the fake news' believability. We therefore hypothesize

H1a. Providing an intervention with warning labels and verified alternative information directly after fake news reduces the fake news's believability *during exposure* compared to providing an intervention before the fake news.

H1b. Providing an intervention with warning labels and verified information directly after fake news reduces the fake news's believability *over time* compared to providing an intervention before the fake news.

Successful interventions against fake news should reduce its believability during exposure and protect against the long-term believability of the fake news. For this purpose, it is necessary to establish a correct mental model. At first glance, intervening after fake news seems counterintuitive, as research has shown that it is hard to correct an existing mental model [38]. Yet research also indicates that the more pronounced a mental model is, the harder it is to correct [73]. On social media, where news articles often consist of aggregated and concise information (e.g., 280 characters for nonsubscription users of X), the time between exposure to fake news and exposure to an intervention is typically a few seconds at most.

We suggest that the time available during initial exposure to fake news is insufficient to build a stable mental model that would require later correction. Instead, the information is processed together with additional verified information to build a mental model. The effect of more intensive System 2 thinking triggered by a discrepancy between fake news and verified information activates higher cognitive resources for conflict resolution. This leads users to engage longer and more intensively with the verified information, enabling them to identify problematic aspects of fake news stories. We therefore propose that the System 2 thinking (triggered by showing verified information after fake news) fosters the adoption of verified information instead of the fake news. As a result, information processed through System 2 thinking contributes to the formation of robust and enduring mental models that can be retrieved more reliably over time. We therefore hypothesize

H2a. Providing an intervention with warnings and verified information after exposure to fake news results in *less integration of fake news* into the mental model than intervening before exposure to fake news.

H2b. Providing an intervention with warnings and verified information after exposure to fake news leads to *more integration of verified information* into the mental model than providing the intervention before exposure to fake news.

Fig. 1 summarizes the hypotheses in relation to the respective underlying theories.

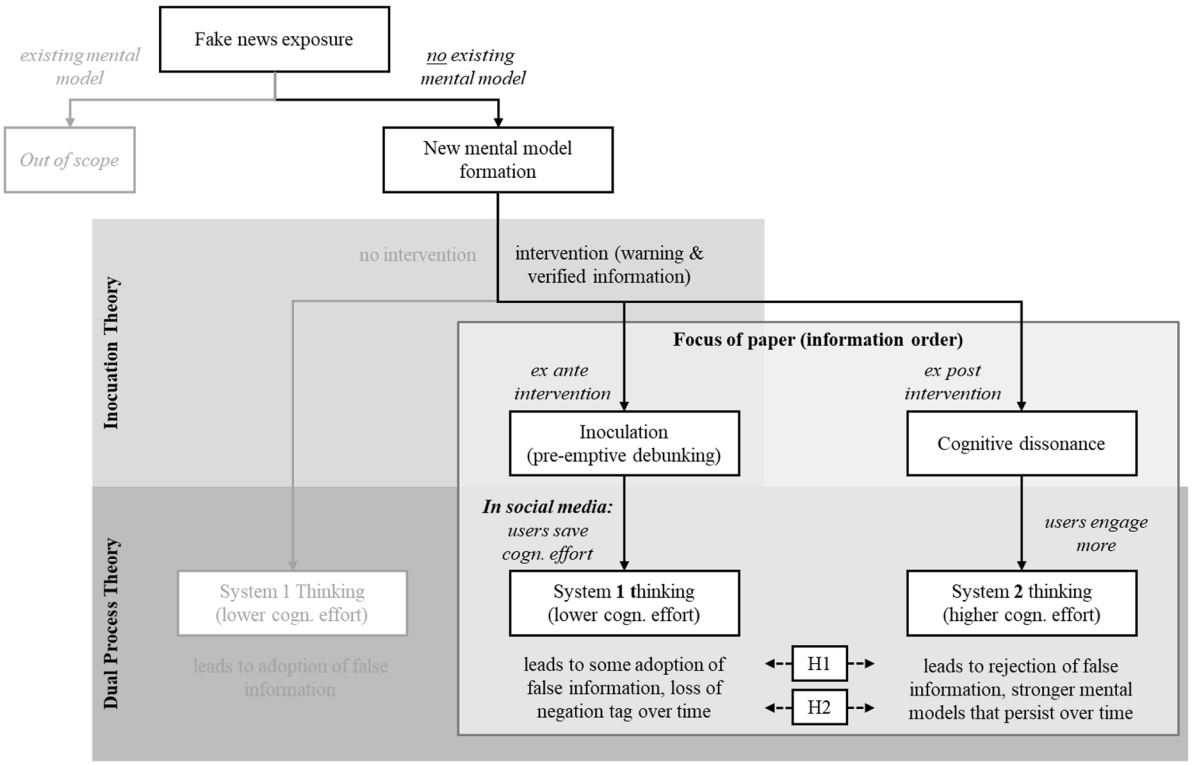


Fig. 1. Schematic presentation of the development of the hypotheses.

4. Method

We conducted a longitudinal online study to investigate our hypotheses in two phases. In phase 1, we exposed the participants to fake news in combination with different interventions and measured short-term effects. Phase 2, which took place two weeks later, examined the effects over time. We conducted a pre-test with nine participants before phase 1. We used the feedback to evaluate the comprehensibility of the instructions, the material, and the procedure, and revised them accordingly.

4.1. The participants

For phase 1, we recruited participants from the United States via Amazon Mechanical Turk (MTurk) in August 2022. We chose MTurk for its broad accessibility and ability to recruit a diverse sample swiftly, ensuring reliable data collection when appropriate screening measures were applied [1]. The participants received $3.50 in monetary compensation. Phase 2 took place two weeks later. The two-week gap was chosen to assess the longer-term effects and the retention of the intervention. This timeframe represents a "stable and generalizable estimate of social life" ([76], p. 287) and has been used in previous IS studies concerned with the long-term impacts of stimuli [10,14]. In contrast, the immediate effects were captured after a five-minute interval in phase 1, consistent with IS studies on short-term reactions to stimuli [75]. The participants received another monetary compensation ($2.00) for their participation.

To ensure high data quality, we selected experienced and reliable workers (based on a worker's approval rating and the number of approved tasks) and implemented several attention checks. First, a question at the start of the survey contained an instruction in the text not to select any answer to the question. If the question was answered, the study was terminated. Second, the participants described how they evaluated news believability in an open-text question. The first author reviewed the participants' responses to filter out invalid or nonsensical responses (e.g., *good study*). We removed three participants owing to invalid responses. Third, we checked which participants were already familiar with the general topics and news stories provided in the survey and were therefore unsuitable as participants (because they already possessed mental models of the subject).

We excluded 23 participants who did not meet the initial criteria and focused on those with no prior knowledge of the general topics relating to the fake news stories. We also removed 60 responses from participants who indicated familiarity with at least one of the presented fake news stories. Ten participants who had never used social media were also excluded. Ultimately, we included only those who participated in both phases 1 and 2. The initial sample consisted of 622 participants. After the attention checks, the final sample comprised 484 participants, with 287 in the main treatments and 197 in the additional control treatments. All participants were from the United States. The sample mostly comprises employees ($n = 307$) and self-employed individuals ($n = 89$), with an average age of 44.60 years. Of the participants, 52 % were female ($n = 254$) and 45 % were male ($n = 218$); 3 % preferred not to disclose ($n = 12$); 56 % ($n = 272$) self-rated politically as rather liberal and 44 % ($n = 212$) as rather conservative. More details are provided in Appendix A.

An institutional review board evaluated the study's ethical aspects and confirmed that the study does not affect the participants' physical or psychological integrity.

4.2. The tasks

In phase 1, we exposed the participants to news stories, of which three were fake. The fake news stories were self-developed to ensure that the participants were unfamiliar with the stories and did not yet have a mental model. We checked whether all the news was new to the participants and excluded individuals who claimed to already know at least one story. After each article, we measured the news's perceived believability. Asking about a story's believability after each article could prime the respondents to be more critical of the next story. To address this, we randomized the order of the news articles. After a filler task

(lasting about 5 min), we asked them questions about the news content and what they believed happened in each story. This step measured the information that entered the participants' mental models. In phase 2, the participants were asked again, after two weeks, what they believed happened in each story, to measure the interventions' impacts on their mental models over time.

4.3. The treatments

We confronted the participants with two fake news interventions and varied their orders to observe differences in the perceived believability and the emerging mental model. To test our hypotheses, we used a between-subjects design with three main treatments:

(1) Control treatment: No intervention.
(2) Ex ante intervention: The participants were first shown the warning label and verified information, and then the fake news story.
(3) Ex post intervention: The participants were first shown the fake news story, and then the warning label and additional information.

We randomized the assignment of participants to the treatments. The number of participants per treatment varied slightly between 92 and 100. To exclude the influence of other variables on the results, we tested for the structural equality of the three treatments. We performed an ANOVA (for age, political orientation, and social media usage) and χ^2 tests (for gender, education, and occupation). At a significance level of 5 %, no differences between the treatments can be observed; thus, we assume structural equality. Appendix B contains the results regarding the structural equality of treatments.

We used two additional control treatments with only warning labels and no verified information to isolate the verified information's influences. Since these effects were not the focus of this work, we considered only the control treatment and the interventions with verified information (i.e., the three main treatments), as follows. (A description of the additional treatments appears in Appendix C.)

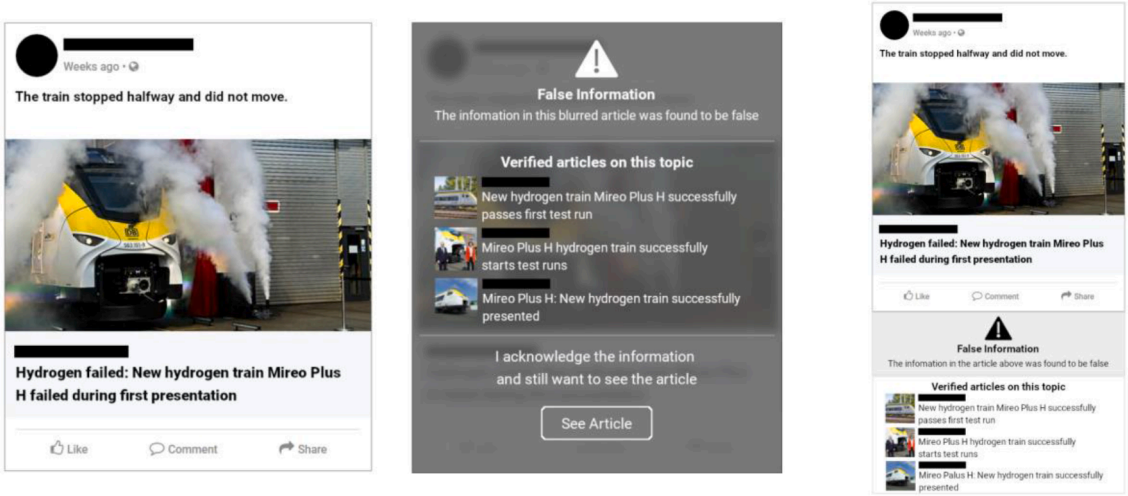
4.4. The material

We developed 21 social media posts. We based the design on the most popular social media platform, Facebook, to increase realism and create a familiar environment. The posts consisted of source, image, headline, and lead sentence. However, as in previous studies [34], we blacked out the source to make the respondents focus on the content. Similarly, we excluded other confounding factors, such as the number of likes and comments.

The social media posts were split into three categories. In social media, fake news does not occur in isolation but between private and real news. We first developed three neutral posts, each containing a short video to simulate private posts from the users' social environment. The videos were of cats (length: 30 s), a sunset (length: 30 s), and a time-lapse of growing plants (length: 19 s). Second, we created 12 news articles with correct and verified information from authentic news sources. The posts of these two categories served as the environment in which we distributed fake news. Third, we developed six fake news articles, two in each of three categories: politically conservative, politically liberal, and without political content. Each participant was randomly shown one conservative, one liberal, and one neutral fake news story. The fake news articles were self-developed to avoid participants' prior knowledge. We also controlled for participants' prior knowledge of all news articles. Appendix D contains all the posts.

In addition to the social media posts, we developed different intervention elements. First, we designed a warning label following the examples of Moravec et al. [49] and Grady et al. [24]. We used a stop icon to draw each participant's attention to the potential threat of fake news and showed a text explaining that the information in the article was false. Thus, we communicated clearly to the participants that this was fake news. Second, we implemented verified information for the events in each fake news article, following Gimpel et al.'s [22] design. Thus, three more news headlines were displayed with an image, again with the sources blacked out. A caption that indicated that the articles contained verified information was also displayed to avoid confusion.

Finally, we varied the order in which the interventions were displayed in accordance with our treatments. For ex ante interventions, we followed Grady et al.'s [24] approach. The fake news article was blurred, and the intervention was displayed as an overlay. Further, to make it clear that the upcoming news article contained false information, the participants had to actively confirm that they had acknowledged the warning label and the verified information and that they still wanted to see the fake news articles. Without this acknowledgment, the participants could not proceed with the study. The fake news article was displayed first for the treatments with ex post interventions, with the warning label and the verified information displayed below it. Fig. 2 contains an overview of the materials in the different treatments.



(1) The control treatment (2) The ex ante intervention (3) The ex post intervention

Fig. 2. Overview of the intervention design in each treatment.

4.5. Independent variables and controls

Prior knowledge and familiarity with topics: To better understand how different intervention styles affect the mental model formation, we first ensured that the participants did not have a mental model by measuring their familiarity with the study topics. We adapted a five-point Likert scale for self-reporting on Internet know-how by Kang et al. to our content topics [31]. Across all measures, we used the original Likert scales in order to remain consistent with prior research. We measured how familiar the participants were with the topics by providing five discrete answers: (1) I've never heard of it; (2) I've heard of this but I don't know what happened or is happening; (3) I know what happened or is happening, but I don't know any specifics; (4) I generally know about it and know some of the specifics; and (5) I am very well informed about it.

Familiarity with the topics: In addition to familiarity with the topics, we measured whether the participants already knew the specific news articles to rule out the presence of an existing mental model. We used Grady et al.'s scale [24] to track the recognition of news articles, with the available answers being *yes* and *no*.

Confirmation bias: To measure confirmation bias, we used the approach suggested by Kim and Dennis [32], which consists of importance and position. We measured the perceived importance of the news article for the participant along a seven-point item (Do you find the issue described in the article important?) (1 = not at all to 7 = extremely) and multiplied it by the participant's position (–3 = extremely negative to +3 = extremely positive). The resulting scores varied between –21 and + 21 [32].

Political orientation: We assessed political orientation using the corresponding scale by Wohl and Branscombe [77] based on an eight-point item (Politically, I would say I am … [1 = very liberal to 8 = very conservative]).

Social media usage: To measure the regularity of social media use, we used the corresponding scale by Gimpel et al. [23] based on a six-point item (How often do you use online social networks (e.g., Facebook, Twitter, or Instagram) in your everyday life? [1 = never to 6 = several times a day]).

4.6. Dependent variables

Believability: To measure the perceived believability of each news story, we used the established seven-point Likert scale by Kim and Dennis [32], which consists of three items.

Mental model: A mental model can be a complex and multifaceted construct. However, we controlled the participants' lack of mental models. To measure which of the information shown during the study participants had in their mental models, we used a simplified recognition test analogous to the one used to study the misinformation effect [69]. To assess mental model formation, we did not merely measure the participants' ability to recall specific information, but also examined how they connected different pieces of information and integrated them into a consistent mental framework.

After exposure to the fake news and the various interventions, we asked the participants to respond to a simulated conversation. The response options were designed to differentiate between mere retention of facts (e.g., *I heard that…*) and the integration of verified knowledge into a coherent narrative. This approach allowed us to analyze whether the participants had simply remembered individual pieces of information, or whether they had constructed a mental representation of the event. By distinguishing between information retention and mental model, we gain deeper insights into the long-term effectiveness of interventions against fake news. To increase realism, we formulated them like questions from another person, simulating a conversation on the topic (e.g., *I heard that a hydrogen train was recently introduced. Do you know how the presentation went?*). We provided five response options, in random order: (1) with the content from the verified information, (2) with the content from the fake news story, (3) and (4) with new fictional content that was not previously shown, and (5) a response that no credible information was received. Thus, we could distinguish whether participants had adopted fake news stories, verified information, or neither into their mental models.

Recognition of the articles: To measure which parts of an article and an intervention were remembered after exposure, we developed a custom five-point Likert scale to capture how well they remembered the article, the warning label, and the verified information (e.g., Do you remember seeing a warning message [during the first session] stating the information in this article was found to be false? [1 = definitely no to 5 = definitely yes]). We report all the scales in Appendix E.

4.7. The procedures

4.7.1. Phase 1

In step 1 in phase 1a, we informed the participants that they were participating in a study on social media behavior. We deliberately did not mention fake news. We then asked them about their previous knowledge of the news articles' topics. They then underwent a tutorial in which we explained the individual components of a social media post. Depending on the treatment, we provided them with an explanation of the included intervention elements. The tutorial sought to ensure that participants understood which information types they would see during the study.

Step 2 was exposure to news. Here, we showed the participants nine news articles in random order: six real news articles and three fake ones. After every third news article, we showed one of the short videos. Depending on the treatment, the posts contained interventions. After seeing each post, we asked the participants about the article's believability and their potential confirmation bias.

Step 3 was a filler task. The participants had to solve a 7 × 4 memory game in under five minutes. To increase their cognitive effort, the goal was to find the corresponding solution to simple mathematical equations (e.g., 25 – 8 and 17). After five minutes, the participants could continue even if they had not finished the task yet. Such filler tasks are used in research to empty participants' short-term memory and to simulate the passage of time [73]. This procedure allowed us to observe the interventions' short-term effects.

Fourth, in phase 1b, we measured the mental model for three real and the three fake news stories. The adjustment from six to three real news stories between the phases was made to reduce the experiment's complexity while ensuring that the primary focus remained on evaluating the interventions' impacts on the fake news over time. While the participants were not incentivized to provide specific or socially desirable responses, they were informed that, during the mental model measurement, they could receive a small bonus of $0.10 per correct answer. This incentive was introduced solely to encourage engagement with the content during the mental model measurement and did not influence the participants' initial responses during exposure to news and other measurement steps. We also ensured that they would not receive the information regarding their achieved individual bonus until the end of phase 2 (regardless of participation), so that they could not draw any conclusions regarding the news articles.

Fifth, through article recognition we checked how the believability had changed and whether the participants could remember the original articles and interventions. For this purpose, we showed them the three fake news articles again and three real news articles in random order. Here, the articles were shown in a reduced form (i.e., without pictures), so that the participants would focus only on the content and would not recognize the pictures. We also added six new real news articles. As in step 2, we asked the participants about each news article's believability. Thus, we asked them if they remembered the exact article, the warning label, and the verified information.

Phase 1 ended with a short survey on demographic characteristics.

4.7.2. Phase 2

We followed Grady et al.'s [24] example, using a two-week period to observe long-term changes in the interventions' effects. We informed the participants via e-mail about possible participation before the start of phase 2. In phase 2, the renewed measurement of the mental model began directly after a brief introduction, followed by the article recognition test. The procedure was identical to that in phase 1. Fig. 3 summarizes the study design.

5. Results

5.1. Descriptive data analysis

An initial descriptive analysis revealed a clear distinction between the treatment groups. Fig. 4 illustrates the mean believability of the fake news per treatment across different points in time for each participant. The intervention groups had a lower average perceived believability of fake news. The descriptive data suggest that this difference is particularly pronounced during exposure to fake news and decreases over time. Regarding the order of interventions, the ex post intervention consistently had the lowest believability. The results show that this effect decreases over time. Appendix F contains a detailed descriptive analysis, while Appendix G contains a descriptive overview of the data.

Fig. 5 visualizes the measurements of the mental models. We considered the number of fake news stories adopted into each participant's mental model, in other words, the number of fake news stories selected by the participants as answers in a conversation on the topic. This yielded a value between zero (no fake news story was adopted into their mental model) and three (all fake news stories were adopted into their mental model). The descriptive results reveal that the participants who had received interventions reported less fake news. Conversely, several control treatment participants more often reported two or three fake news stories. Further, the ex post intervention treatment shows the best results.

Comparing these results with phase 2, we see that the difference became smaller, but we observe the same order. In sum, examining how

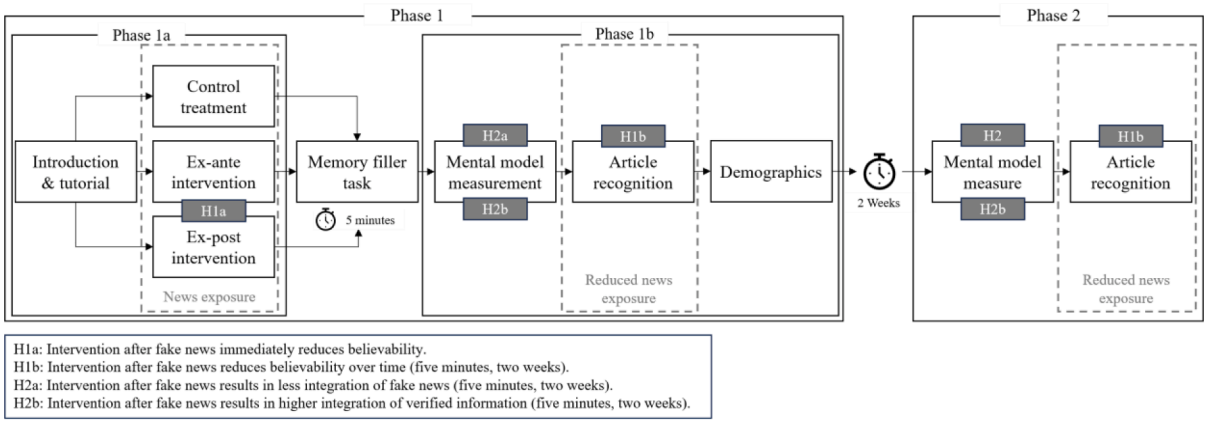
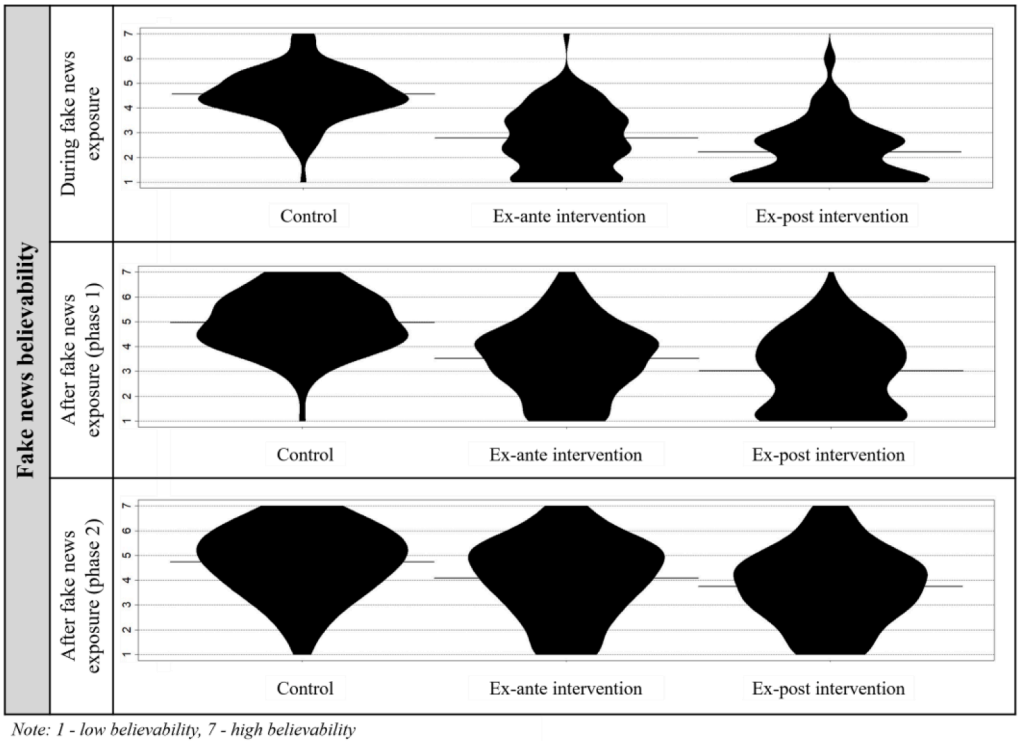


Fig. 3. Summary of the procedure of the study.



Note: 1 - low believability, 7 - high believability

Fig. 4. Comparisons of the mean believability of the fake news at different points in time. The mean value of the distribution is shown as a horizontal line.

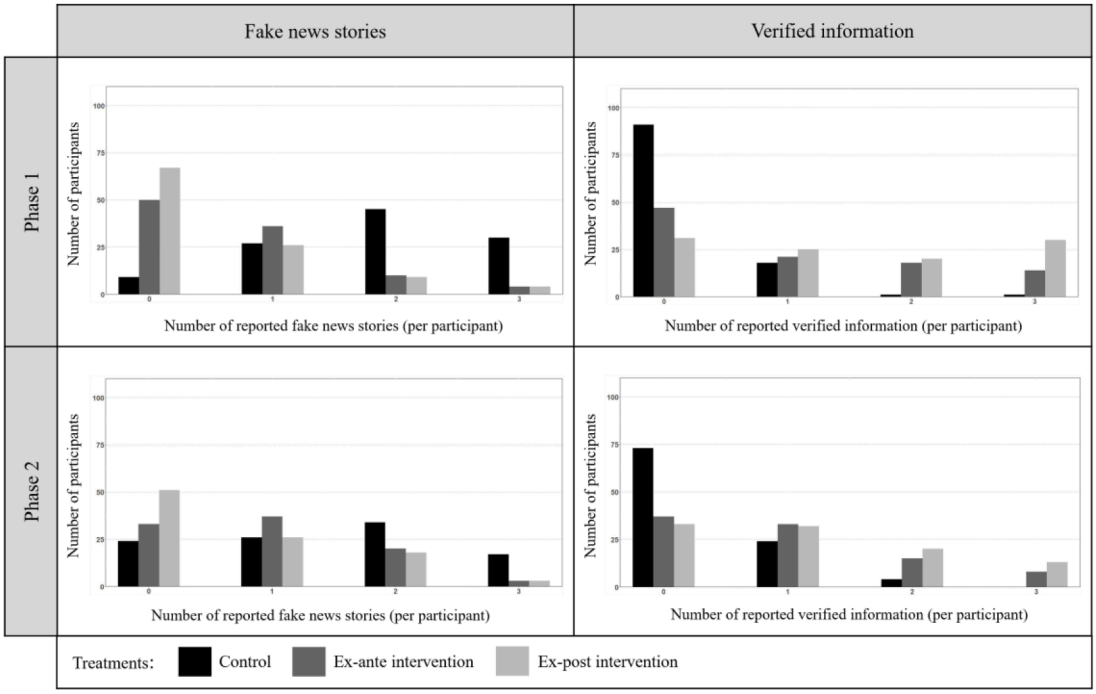


Fig. 5. Overview over the mental model measurement per treatment at different points in time. Numbers vary from zero (the participant adapted no fake news story/verified information into their mental model) to three (they adapted all fake news stories/verified information into their mental model).

many participants have taken the verified information into their mental models, we see, in phases 1 and 2, that the ex post intervention treatment yielded the highest reporting of the verified information.

5.2. The testing of the hypotheses

5.2.1. The effects of the order of the interventions on believability

To test the effects of intervention order during exposure to fake news (H1a) and after such exposure (H1b) on believability, we compared the different treatments along believability. The Shapiro–Wilk test and the Anderson–Darling test [61] suggest that the data were not normally distributed. We therefore used a one-sided nonparametric Wilcoxon–Mann–Whitney pairwise comparison. Table 1 contains an overview of the results, while Appendix H contains a detailed report of the results.

To test H1a, we compared the average believability of the three fake news articles during exposure. The average believability with ex post intervention was lower than with ex ante intervention and in the control treatment. Compared with the control treatment, both the ex post intervention (p-value < 0.001) and the ex ante intervention (p-value < 0.001) show a significant positive effect. Further, we see a much stronger positive effect on believability of the ex post intervention than of the ex ante intervention (p-value < 0.001), supporting H1a.

To examine the intervention orders' effects over time, we look at the average perceived believability in phases 1b and 2. Phase 1b's results show that both the ex post and the ex ante interventions led to significantly lower believability (p-value < 0.001 for both tests) than the control treatment. At the same time, the data show that the ex post intervention had a significantly larger positive effect on the fake news's believability than the ex ante intervention (p-value $= 0.005$). This effect decreased over time. Phase 2′s results still show a reduction in believability in the ex post and the ex ante interventions compared with the control treatment (p-value $= 0.002$ and p-value < 0.001). Two weeks after the exposure, we also observe a significant difference between the intervention types (p-value $= 0.049$), confirming H1b.

5.2.2. The effects of intervention order on the formation of mental models

To test the effects of the order of intervention on the mental models, first, we compared how many fake news stories were adopted by the

Table 1
Results of the one-sided Wilcoxon-Mann-Whitney pairwise believability comparisons.

Believability		Mean	Standard deviation	p-values (Wilcoxon–Mann–Whitney)				
				Control	Ex ante intervention		Ex post intervention	
Phase 1a	Control	4.569	1.024	—	< 0.001	***	< 0.001	***
(H1a)	Ex ante intervention	2.785	1.280		—		< 0.001	***
	Ex post intervention	2.220	1.176				—	
Phase 1b	Control	4.972	1.174	—	< 0.001	***	< 0.001	***
(H1b)	Ex ante intervention	3.512	1.469		—		0.019	*
	Ex post intervention	3.022	1.509				—	
Phase 2	Control	4.743	1.404	—	0.002	**	< 0.001	***
(H1b)	Ex ante intervention	4.050	1.604		—		0.049	*
	Ex post intervention	3.766	1.521				—	

$+ p < 0.1$.
$^{*} p < 0.05$.
$^{**} p < 0.01$.
$^{***} p < 0.001$.

Table 2
The results of the one-sided Wilcoxon–Mann–Whitney test comparing the fake news adoption into the mental model.

Number of fake news articles adopted into their mental models		Mean	Standard deviation	*p*-values (Wilcoxon–Mann–Whitney)				
				Control	Ex ante intervention		Ex post intervention	
Phase 1b	Control	1.889	0.891	—	< 0.001	***	< 0.001	***
(H2a)	Ex ante intervention	0.696	0.835		—		0.102	
	Ex post intervention	0.573	0.830				—	
Phase 2	Control	1.434	1.042	—	< 0.001	***	< 0.001	***
(H2a)	Ex ante intervention	0.925	0.837		—		0.017	*
	Ex post intervention	0.725	0.871				—	

+ $p < 0.1$.
* $p < 0.05$
** $p < 0.01$.
*** $p < 0.001$.

Table 3
Results of the one-sided Wilcoxon–Mann–Whitney test comparing the adoption of the verified information into the mental models.

Number of verified information adopted into their mental models		Mean	Standard deviation	*p*-values (Wilcoxon–Mann–Whitney)				
				Control	Ex ante intervention		Ex post intervention	
Phase 1b	Control	0.192	0.420	—	< 0.001	***	< 0.001	***
(H2b)	Ex ante intervention	1.000	1.119		—		0.012	*
	Ex post intervention	1.375	1.172				—	
Phase 2	Control	0.313	0.547	—	< 0.001	***	< 0.001	***
(H2b)	Ex ante intervention	0.935	0.953		—		0.069	+
	Ex post intervention	1.133	1.032				—	

Notes: The control treatment contained no verified information.
+ $p < 0.1$.
* $p < 0.05$
** $p < 0.01$.
*** $p < 0.001$.

participants into their mental models. Second, we compared how many verified information stories were adopted. The Shapiro–Wilk test and the Anderson–Darling test [61] revealed that the data were not normally distributed. We therefore used a one-sided nonparametric Wilcoxon–Mann–Whitney pairwise comparison. Tables 2 and 3 summarize the results.

To test the ex post intervention's effects on fake news integration into the mental models, we considered the number of fake news stories adopted by the participants into their mental models in phase 1b The evaluation of the Wilcoxon–Mann–Whitney pairwise comparison with the control treatment shows that both treatments led to less fake news in a participant's mental model (p-value < 0.001 for both tests). However, a direct comparison of the intervention types shows no significant short-term reduction of fake news in the mental model (p-value $= 0.102$).

Further, we considered the results from two weeks later (phase 2). Both interventions continued to have significant positive effects (p-value < 0.001 for both tests) compared with the control treatment. A direct comparison of the intervention types shows a significant positive effect of the ex post intervention on the integration of fake news stories into the mental models over time (p-value $= 0.017$), partially supporting H2a.

To examine the effects of the ex post intervention on the adoption of the verified information into the mental model, we considered the number of verified information stories recalled in the mental models. We observe in phase 1b that participants in the ex post intervention treatment were able to recall the verified information in the mental model more often than in the ex ante intervention treatment. Thus, a Wilcoxon–-Mann–Whitney pairwise comparison shows a statistically significant difference (p-value $= 0.012$). However, the effect decreased over time. In phase 2, we see more verified information adopted into the mental model in the ex post intervention treatment compared to the ex ante intervention, but the effect was not statistically significant (p-value $= 0.069$), partially supporting H2b. Table 4 summarizes all the empirical results.

To complement our analysis, we conducted an exploratory examination of demographic influences on the perception of fake news (see Appendix I). While gender, age, occupation, and education showed no significant effects, political orientation was consistently associated with higher believability ratings, and social media usage showed a significant positive relationship in the later study phase. These findings suggest that individual differences, particularly political orientation and social media usage, may influence how fake news is perceived. However, as this was not our primary focus, these results should be interpreted with caution. The additional analysis (detailed in Appendix I) provides further context, but does not establish causal relationships. Researchers are encouraged to explore these factors in greater depth to better understand their role in the processing of misinformation. A supplementary analysis of potential outliers, which does not affect the results, can be found in Appendix J.

Table 4
Overview of the hypotheses and the empirical results.

Theoretical hypotheses		Empirical results
H1a	Ex post intervention reduces the fake news' believability compared to ex ante intervention (during exposure)	Supported
H1b	Ex post intervention reduces the fake news' believability compared to ex ante intervention (over time)	Supported
H2a	Ex post intervention reduces the integration of fake news into the mental models compared to ex ante intervention (over time)	Partially supported
H2b	Ex post intervention increases the integration of verified information into the mental models compared to ex ante intervention (over time)	Partially supported

6. Discussion

The empirical results confirmed our hypothesis that intervention with a warning message and verified information after the exposure to

fake news (ex post) results in lower believability than the same intervention before the exposure (ex ante), even though the same information is displayed. This is consistent with the findings of Pal et al. [58], who also observed that the timing of counterarguments influences how individuals process and accept rebuttals. Specifically, they found that counterarguments presented after a rumor – analogous to our ex post intervention – are more effective in changing perceptions of credibility and persuasiveness.

Per our hypothesis, respondents engaged more deeply with the information when attempting to correct the false information they had initially encountered. This aligns with the findings of Moravec et al. [51], who showed that cognitive dissonance – triggered by conflicting information – drives users process and correct disinformation more thoroughly. This deeper engagement with the validated information facilitates its integration into mental models, consistent with Kahneman's [30] System 2 thinking. Recent research on fake news further supports this mechanism, showing that interventions that promote deeper cognitive engagement – beyond superficial interaction – can help users reconsider and revise false beliefs [70]. Our results extend these findings by showing sustained effects over time. Deeper processing, driven by the need to resolve the conflict between false and true information, helps to integrate verified information more effectively into users' mental models while reducing the incorporation of fake news. Thus, providing corrective information after exposure to false content leads to more robust and lasting updates to mental models. This supports the theoretical perspective that effective interventions must consider the sequence of information presentation to maximize cognitive engagement and the correction of disinformation [35,47].

When the participants first received the fake news articles, they were more likely to resolve the conflict created by the following verified information by activating System 2 thinking and recognizing the problematic aspects of the fake news stories. This process is analogous to how Zúñiga et al. [80] described user elaboration as facilitating deeper cognitive processing and increased corrective action. Besides the positive effect on believability during exposure to fake news, we also observed a lasting effect over time. Our observation of a lasting effect of interventions over time is notable, because it contrasts with the findings of Grady et al. [24], who reported that the impacts of both ex ante and ex post interventions when a single warning message was used diminished over time and that no significant difference remained between the two intervention types.

We found evidence that an intervention administered after exposure to fake news (ex post) leads to reduced adoption of the fake news stories into users' mental models. However, this effect only becomes visible over time. After two weeks, we see more clearly that participants who received the ex post intervention integrated significantly less fake news into their mental models than those who received the intervention beforehand (ex ante). This delayed effect is expected, as under both intervention conditions (ex post and ex ante) the participants had a clearer short-term memory of what they had seen (e.g., the warning labels) five minutes after the exposure to fake news – resulting in smaller initial differences. Yet, over time, the ex post intervention led to a lower incorporation of fake news into the participants' mental models and a stronger integration of verified information. This suggests that the cognitive dissonance triggered by conflicting information promotes deeper cognitive processing [47,51]. In contrast, the ex ante intervention appears to primarily support the retention of warning labels without fostering the same depth of cognitive engagement. These findings reinforce the view that mental model formation does not occur merely through exposure to information, but requires active cognitive reconciliation between previous and new knowledge. Over time, this contributes to a more accurate mental representation and a reduction in the retention of fake news.

Second, the participants adopted more verified information into their mental models after exposure to fake news when the intervention occurred after the fake news story. Because the participants engaged with the information more intensively when the intervention occurred after the exposure to fake news, they could recall the correct information more easily later [8]. However, this effect weakened over time. Although we observed a positive effect, it was no longer significant after two weeks. Admittedly, the challenge for the participants in phase 2 was high. They had to recall externally provided information that they had no personal connection with, and that was shown to them two weeks earlier as part of a single social media post. Thus, it seems plausible that the gap narrows over time without repeated exposure. This interpretation is consistent with recent findings that deeper cognitive engagement is necessary to support retention [70].

Third, owing to the differences in the mental models, we also see reduced believability of fake news stories. This effect was also significant after two weeks. It may seem counterintuitive that we see no significant differences in perceived believability, although the participants adopted less fake news into their mental models. This phenomenon aligns with our expectations that the cognitive dissonance from first receiving fake news nudges users to invest cognitive effort [51]. This behavior helps them identify the problematic fake news content and reduce the adoption of false information into their mental models, reducing the fake news' believability during exposure and over time.

Fourth, to examine the interventions' effects on the perception of real news, we also investigated whether there were differences in the real news's believability between the treatment groups and the control group. This allowed us to analyze whether the interventions against fake news had unintended effects on the perception of real news. However, the Kruskal–Wallis test for phases 1a, 1b, and 2 revealed no significant differences in the real news's believability between the groups. Thus, the interventions specifically influenced the perception of fake news without affecting the real news's believability. Appendix K contains detailed information.

Our findings suggest that well-designed interventions can reduce disinformation's impacts without fostering skepticism toward real news. Previous studies indicate that, in the absence of source cues, clarity, coherence, and logical structure shape credibility judgments [34,38]. Cognitive fluency (i.e., the ease of processing information) also plays a key role [60]. By improving the presentation and structure of verified content, future interventions could better facilitate cognitive fluency and engage System 2 processing, making it easier for individuals to process and accept accurate information. This is particularly relevant in cases where real news is initially rejected because it contradicts an individual's existing beliefs. When encountering such news, individuals often avoid the cognitive effort required to resolve confirmation bias, leading them to dismiss the information rather than integrate it into their worldview. However, interventions that encourage deeper analytical processing—such as prompting users to actively compare conflicting pieces of information—could help lower this resistance. While we did not directly test this, the literature suggests that fostering analytical thinking can reduce resistance to counterintuitive information [60]. Thus, interventions that promote deeper processing may both combat fake news and encourage a more reflective approach to the consumption of real news.

In sum, the order in which an intervention is presented impacts the perception of fake news. The use of an intervention directly after exposure to fake news leads to the participants believing the fake news less, adopting it less into their mental models, and continuing to believe it less after the exposure without influencing the perception of real news. A comparison using only a warning label is provided in Appendix L.

6.1. *Contributions to theory*

The current body of knowledge on interventions against fake news follows two main theories, which we have incorporated into our study (see Fig. 1). Inoculation theory suggests that individuals should be forewarned against potential fake news and that this fake news should be pre-emptively debunked [39] . Yet research shows that negation tags

can be lost over time, limiting long-term effects. Also, recent IS research has utilized dual process theory to propose that active cognitive engagement (System 2 thinking) with fake news can mitigate its persuasiveness [53] and can increase interventions' effectiveness [49]. We have extended this knowledge base as follows.

First, we have provided empirical support to show that the order of information can influence an intervention's effectiveness against fake news and the perceptions of fake news, particularly over time. Thus, we have challenged the conventional wisdom relating to inoculation theory that calls for pre-emptive debunking. By investigating over two weeks, which "represents a stable and generalizable estimate of social life" ([76], p. 287), we have contributed longitudinal empirical evidence for the design and study of interventions [24]. This extends existing knowledge by demonstrating that interventions' content and sequence are critical. Our study goes beyond previous work by showing how different intervention sequences affect the durability of mental models, providing a deeper understanding of long-term intervention effectiveness. The findings highlight the critical role of information order in designing effective interventions, offering a nuanced understanding of how interventions can leverage cognitive processes to reduce disinformation [35,80].

Second, we have provided a novel explanation for the cognitive process involved in exposure to fake news and interventions, contributing knowledge regarding dual process theory. In particular, we have provided insights into how individuals form mental models, what influences intervention order has, and what the consequences are for perceived believability over time. We have shown that presenting interventions after exposure to fake news triggers System 2 thinking and affects the formation and stability of mental models; we suggest that this is due to the cognitive dissonance that is created. Our findings are in line with the research into cognitive elaboration [80], which showed that interventions that enhance cognitive engagement can lower the spreading of fake news. Our findings and conclusions may explain why warning interventions alone are insufficient for long-term effectiveness [51,64]. Instead, we have shown that fostering robust mental models through a combination of warnings and verified information in the right order is crucial to effectively counter fake news over time.

Our findings on the cognitive processes involved in perceiving fake news and the influences of intervention orders add to the emerging body of IS research on measures against fake news. Thus, we have responded to the call for more research into the cognitive processes involved in dealing with fake news and the need for longitudinal empirical studies [24]. Our findings can serve as a starting point for developing new interventions as well as improving existing approaches.

6.2. Practical implications

Our findings can help practitioners and social media platform providers to develop interventions against fake news. First, warning labels alone are insufficient to achieve an enduring effect in the fight against fake news. In addition to a warning, verified information should be provided to users to help them build a mental model of the situation. Second, we encourage developers to incorporate mental processes into decision-making when designing interventions against fake news. In fast-paced, information-dense environments, users try to save cognitive resources through System 1 processing. Thus, platform owners should incentivize System 2 engagement for potentially false information.

We have shown that additional information should be displayed after fake news to motivate users to invest more cognitive resources and engage more with the content. Today, platforms such as X and Instagram flag such content in different ways, for instance, through community notes, warning labels, or fact-checking results. Our findings could inform the design of these platform features. For instance, platforms can implement a delayed engagement mechanism that introduces a waiting period of a few seconds before users can share or like disputed content. Further, research has found that dialogues with chatbots can mitigate fake news [16]. Platform providers could implement chat popups through which users can discuss disputed news items. Third, in line with previous research [24], we observed lower intervention effectiveness. Thus, interventions should be used regularly to support social media users. While this may be a difficult undertaking, platforms could individually curate fact-checked news using machine learning or encourage re-exposure to personalized fact-checking reminders in a user's feed. As pointed out in previous work, it is unlikely that all users will choose to engage with such interventions [16]; these authors therefore suggest incentive programs and gamification as future applied work. However, we agree with the authors that practical challenges – such as scalability, user adoption, and user fatigue from frequent interventions – remain unresolved.

6.3. Limitations and future research

Our results have limitations. The design of our study materials is one. Our fake news and verified information were designed to show clear contrast. Yet the line between fake news and true information is often not so obvious, and verified information is not always available. In real-world settings, the differences may be smaller than those we observed. Future research could verify this in real-world settings.

Further, there were subtle differences between the treatments' designs. In the treatment with the interventions before the fake news (ex ante), the participants had to read and actively confirm the intervention information and only saw the fake news after confirming. The intervention appeared directly below the fake news article in the ex post treatment. We cannot eliminate the possibility that this simultaneous presentation influenced the results, as the participants were able to compare the differences more carefully.

The study duration also limits the generalizability of the results over time. The measurements were carried out approximately five minutes and two weeks after exposure to the fake news. In phase 1, this allowed us to simulate the critical moments when fake news begins to circulate. In phase 2, we simulated the situation of users re-encountering fake news after a longer period. However, we cannot say whether the effects persisted after more than two weeks; yet two weeks is generally considered an adequate timeframe for such studies [24]. Further research could validate the longevity of our findings and could examine a more extended period. Our results showed that interventions' effects declined over the two weeks. Yet it is unclear whether the effects continue to decline over time, dissipate entirely, or reach a plateau.

A mental model is a complex construct. We have focused on the creation of mental models that do not yet exist. In this context, we analyzed information processing. However, participants may still have pre-existing biases or general knowledge about broader topics, even if they were unfamiliar with a specific news story. For instance, many people are likely aware of concepts such as autonomous vehicles, and this pre-existing knowledge may have influenced their responses. Although we controlled for the news stories in our design, biases cannot be wholly eliminated. Nonetheless, this limitation also reflects the realistic nature of information processing in social media. In everyday life, users may also encounter fake news about a topic for which a mental model already exists. For instance, research has shown that phenomena such as confirmation bias [32] make it hard to correct an existing mental model [38]. Future research could shed light on how users process information that conflicts with existing mental models to better understand how fake news works.

Social media users consume information hedonically. While increasing realism is certainly desirable in experimental designs, we deliberately chose to conduct our experiment in a controlled setting at the expense of this desirable real-world condition. Yet our experiment's setting may naturally make participants more focused, which differs from typical social media consumption. We assume that this may lead participants to rely more on System 2 thinking (deliberate with cognitive effort) than in real-world settings. This could lead to the

interventions' effects being underestimated compared to the effects for the control group. Further, it could lead to an underestimation of the difference between ex ante and ex post warning labels – the latter aim to increase System 2 thinking. We are confident that these two effects did not undermine our study findings.

It may be valuable to investigate whether interventions that induce cognitive dissonance and promote System 2 thinking can also counteract confirmation bias in the evaluation of real news. Real news that contradicts an individual's worldview is often rejected automatically, not because of its veracity, but owing to ingrained cognitive biases. Interventions designed to trigger cognitive dissonance may encourage more deliberate processing, increasing the acceptance of real news that initially appears counterintuitive. This approach could serve as a promising strategy to mitigate confirmation bias, ultimately fostering more balanced information processing, even in contexts where verifiable news is typically discounted owing to pre-existing beliefs.

To further advance the understanding of how interventions against fake news work, future research could explore additional dimensions beyond believability, mental model integration, and article recognition. For instance, integrating shareability as a dependent variable may yield valuable insights. Shareability measures the likelihood that users will share fake news articles – a critical factor given social media's roles in amplifying disinformation. By examining how interventions impact shareability, future research could assess whether reducing believability and improving article recognition also lowers the likelihood of propagation.

7. Conclusions

We have investigated the influences of intervention order in combating fake news, emphasizing the importance of understanding how users process information to ensure long-term effectiveness. Based on theoretical reasoning, we derived hypotheses and validated them through a longitudinal study. Our findings demonstrate that altering the sequence – whether the fake news story or the intervention is presented first – significantly impacts an intervention's success. Specifically, presenting additional information after the fake news stimulates cognitive dissonance, prompting System 2 processing, which requires greater cognitive effort and fosters the development of clearer, more accurate mental models. This study advances the understanding of information processing in the context of disinformation and its countermeasures. Moreover, our results offer practical implications for reducing the adverse impacts of fake news on individuals, organizations, and society.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Microsoft CoPilot to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

CRediT authorship contribution statement

Christian Olenberger: Writing – original draft, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Manfred Schoch:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Lena Utz:** Writing – review & editing, Writing – original draft, Validation.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.im.2025.104256.

References

[1] Herman Aguinis, Isabel Villamor, Ravi S. Ramani, MTurk research: review and recommendations, J. Manage 47 (4) (2021) 823–837, https://doi.org/10.1177/0149206320969787.

[2] Esma Aïmeur, Sabrine Amri, Gilles Brassard, Fake news, disinformation and misinformation in social media: a review, Soc. Netw. Anal. Min. 13 (1) (2023) 30, https://doi.org/10.1007/s13278-023-01028-5.

[3] Hunt Allcott, Matthew Gentzkow, Social media and fake news in the 2016 election, J. Econ. Perspect. 31 (2) (2017) 211–236, https://doi.org/10.1257/jep.31.2.211.

[4] Anat Alon, Ilan Rahimi, Hila Tahar, Fighting fake news on social media: a comparative evaluation of digital literacy interventions, Curr. Psychol. 43 (2024) 1–19, https://doi.org/10.1007/s12144-024-05668-4.

[5] F. Arendt, M. Haim, J. Beck, Fake news, warnings and perceptual truth: on the differential sucseptibility for fake news depending on recipients' political orientations, Publizistik (64) (2019) 181–204.

[6] Inaiya Armeen, Ross Niswanger, Chuan Tian (Annie), Combating fake news using implementation intentions, Inf. Syst. Front. 27 (3) (2025) 1107–1120, https://doi.org/10.1007/s10796-024-10502-0.

[7] Michael S. Ayers, Lynne M. Reder, A theoretical review of the misinformation effect: predictions from an activation-based memory model, Psychon. Bull. Rev. 5 (1) (1998) 1–21, https://doi.org/10.3758/BF03209454.

[8] Leif Azzopardi, Cognitive Biases in Search, in: Falk Scholer, Paul Thomas, David Elsweiler, Hideo Joho, Noriko Kando, Catherine Smith (Hg.) (Eds.), Proceedings of the 2021 Conference on Human Information Interaction and Retrieval. Conference on Human Information Interaction and Retrieval, Canberra, Australia, 2021, pp. 27–37.

[9] Sven Beisecker, Christian Schlereth, Sebastian Hein, Shades of fake news: how fallacies influence consumers' perception, Eur. J. Inf. Syst. (2022) 1–20, https://doi.org/10.1080/0960085X.2022.2110000.

[10] Alexander Benlian, A daily field investigation of technology-driven stress spillovers from work to home, Manage. Inf. Syst. Q. 44 (3) (2020), https://doi.org/10.25300/MISQ/2020/14911.

[11] Nadia M. Brashier, Gordon Pennycook, Adam J. Berinsky, David G. Rand, Timing matters when correcting fake news, Proc. Natl. Acad. Sci. U.S.A. 118 (5) (2021), https://doi.org/10.1073/pnas.2020043118.

[12] Kathryn A. Braun, Elizabeth F. Loftus, Advertising's misinformation effect, Appl. Cogn. Psychol. 12 (6) (1998) 569–591, https://doi.org/10.1002/(SICI)1099-0720(1998120)12:6<569::AID-ACP539>3.0.CO;2-E.

[13] Grégoire Burel, Tracie Farrell, Harith Alani, Demographics and topics impact on the co-spread of COVID-19 misinformation and fact-checks on Twitter, Inf. Process. Manage 58 (6) (2021) 102732, https://doi.org/10.1016/j.ipm.2021.102732.

[14] Andrew Burton-Jones, Detmar W. Straub, Reconceptualizing system usage: an approach and empirical test, Inf. Syst. Res. 17 (3) (2006) 228–246, https://doi.org/10.1287/isre.1060.0096.

[15] William E. Crozier, Deryn Strange, Correcting the misinformation effect, Appl. Cogn. Psychol. 33 (4) (2018) 585–595, https://doi.org/10.1002/acp.3499.

[16] T.H. Costello, G. Pennycook, D.G. Rand, Durably reducing conspiracy beliefs through dialogues with AI, Science (1979) 385 (6714) (2024), https://doi.org/10.1126/science.adq1814.

[17] Della Vedova, L. Marco, Eugenio Tacchini, Stefano Moret, Gabriele Ballarin, Massimo DiPierro, Luca de Alfaro, Automatic online fake news detection combining content and social signals, in: 22nd Conference of Open Innovations Association. 22nd Conference of Open Innovations Association, Jyvaskyla, Finland, 2018, pp. 272–279.

[18] Bingjie Deng, Michael Chau, The effect of the expressed anger and sadness on online news believability, J. Manage. Inf. Syst. 38 (4) (2021) 959–988, https://doi.org/10.1080/07421222.2021.1990607.

[19] Aaron M. French, Veda C. Storey, Linda Wallace, The impact of cognitive biases on the believability of fake news, Eur. J. Inf. Syst. (2023) 1–22, https://doi.org/10.1080/0960085X.2023.2272608.

[20] R.Kelly Garrett, Shannon Poulsen, Flagging facebook falsehoods: self-identified humor warnings outperform fact checker and peer warnings, J. Comput.-Mediated Commun. 24 (5) (2019) 240–258, https://doi.org/10.1093/jcmc/zmz012.

[21] Jordana George, Natalie Gerhart, Russell Torres, Uncovering the truth about fake news: a research model grounded in multi-disciplinary literature, J. Manage. Inf. Syst. 38 (4) (2021) 1067–1094, https://doi.org/10.1080/07421222.2021.1990608.

[22] Henner Gimpel, Sebastian Heger, Julia Kasper, Ricarda Schäfer, The power of related articles – improving fake news detection on social media platforms, in: Proceedings of the 53rd Hawaii International Conference on System Sciences. Hawaii International Conference on System Sciences, Wailea Maui, USA, 2020.

[23] Henner Gimpel, Sebastian Heger, Christian Olenberger, Lena Utz, The effectiveness of social norms in fighting fake news on social media, J. Manage. Inf. Syst. 38 (1) (2021) 196–221, https://doi.org/10.1080/07421222.2021.1870389.

[24] Rebecca Hofstein Grady, Peter H. Ditto, Elizabeth F Loftus, Nevertheless, partisanship persisted: fake news warnings help briefly, but bias returns with time, Cogn. Res. Princ. Implic. 6 (52) (2021), https://doi.org/10.1186/s41235-021-00315-z.

[25] Manjul Gupta, Denis Dennehy, Carlos M. Parra, Matti Mäntymäki, Yogesh K Dwivedi, Fake news believability: the effects of political beliefs and espoused cultural values, Inf. Manage. 60 (2) (2023) 103745, https://doi.org/10.1016/j.im.2022.103745.

[26] Katrin Hartwig, Tom Biselli, Franziska Schneider, Christian Reuter, From adolescents' eyes: assessing an indicator-based intervention to combat misinformation on TikTok, in: Proceedings of the 2024 CHI Conference on Human

Factors in Computing Systems, 2024, pp. 1–20, https://doi.org/10.1145/3613904.3642264.
[27] Katrin Hartwig, Frederic Doell, Christian Reuter, The landscape of user-centered misinformation interventions – a systematic literature review, ACM. Comput. Surv. 56 (11) (2024) 1–36, https://doi.org/10.1145/3674724.
[28] Christy Galletta Horner, Dennis Galletta, Jennifer Crawford, Abhijeet Shirsat, Emotions: the unexplored fuel of fake news on social media, J. Manage. Inf. Syst. 38 (4) (2021) 1039–1066, https://doi.org/10.1080/07421222.2021.1990610.
[29] Philip Nicholas Johnson-Laird, Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness, Harvard University Press, Boston, USA, 1983.
[30] Daniel Kahneman, Thinking, Fast and Slow, Macmillan, New York, USA, 2011.
[31] Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, Sara Kiesler, 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. Eleventh Symposium On Usable Privacy and Security, Ottawa, Canada, 2015, pp. 39–52.
[32] Antino Kim, Alan R. Dennis, Says Who? The effects of presentation format and source rating on fake news in social media, Manage. Inf. Syst. Q. 43 (3) (2019) 1025–1039, https://doi.org/10.25300/MISQ/2019/15188.
[33] Hyejin Kim, Euna Mehnaz Khan, Jisu Huh, Jaideep Srivastava, Fighting negative rumors with trust-based rumor-refutation strategy, J. Market. Commun. 29 (7) (2023) 676–700, https://doi.org/10.1080/13527266.2022.2066155.
[34] Antino Kim, Patricia L. Moravec, Alan R. Dennis, Combating fake news on social media with source ratings: the effects of user and expert reputation ratings, J. Manage. Inf. Syst. 36 (3) (2019) 931–968, https://doi.org/10.1080/07421222.2019.1628921.
[35] Bogoan Kim, Aiping Xiong, Dongwon Lee, Kyungsik Han, A systematic review on fake news research through the lens of news creation and consumption: research efforts, challenges, and future directions, PLoS. One 16 (12) (2021), https://doi.org/10.1371/journal.pone.0260080.
[36] Annie Lang, The limited capacity model of mediated message processing, J. Commun. 50 (1) (2000) 46–70, https://doi.org/10.1111/j.1460-2466.2000.tb02833.x.
[37] David M.J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, et al., The science of fake news, Science (1979) 359 (6380) (2018) 1094–1096, https://doi.org/10.1126/science.aao2998.
[38] Stephan Lewandowsky, Ullrich K.H. Ecker, Colleen M. Seifert, Norbert Schwarz, John Cook, Misinformation and its correction: continued influence and successful debiasing, Psychol. Sci. Public Interest 13 (3) (2012) 106–131, https://doi.org/10.1177/1529100612451018.
[39] Stephan Lewandowsky, Sander van der Linden, Countering misinformation and fake news through inoculation and prebunking, Eur. Rev. Soc. Psychol. 32 (2) (2021) 348–384, https://doi.org/10.1080/10463283.2021.1876983.
[40] Elizabeth F. Loftus, Planting misinformation in the human mind: a 30-year investigation of the malleability of memory, Learn. Memory 12 (4) (2005) 361–366, https://doi.org/10.1101/lm.94705.
[41] Bernhard Lutz, Marc T.P. Adam, Stefan Feuerriegel, Nicolas Pröllochs, Dirk Neumann, Affective information processing of fake news: evidence from NeuroIS, Eur. J. Inf. Syst. (2023) 1–20, https://doi.org/10.1080/0960085X.2023.2224973.
[42] Michele Maasberg, Emanuel Ayaburi, Charles Z. Liu, Yoris A Au, Exploring the propagation of fake cyber news: an experimental approach, in: Tung Bui (Hg.) (Ed.), Proceedings of the 51st Hawaii International Conference on System Sciences. 51st Hawaii International Conference on System Sciences, Waikoloa Village, USA, 2018 (Proceedings of the Annual Hawaii International Conference on System Sciences).
[43] Drew B. Margolin, Aniko Hannak, Ingmar Weber, Political fact-checking on Twitter: when do corrections have an effect? Polit. Commun. 35 (2) (2018) 196–219, https://doi.org/10.1080/10584609.2017.1334018.
[44] Miriam J. Metzger, Andrew J. Flanagin, Credibility and trust of information in online environments: the use of cognitive heuristics, J. Pragmat. 59 (2013) 210–220, https://doi.org/10.1016/j.pragma.2013.07.012.
[45] Stacy Miller, Philip Menard, David Bourrie, I'm not fluent: how linguistic fluency, new media literacy, and personality traits influence fake news engagement behavior on social media, Inf. Manage. 61 (2) (2024) 103912, https://doi.org/10.1016/j.im.2023.103912.
[46] Dan Milmo, Jennifer Rankin, EU raises prospect of big fine or ban if twitter fails to follow new legislation. Hg. v, The Guardian (2022) Online available under https://www.theguardian.com/technology/2022/nov/30/eu-raises-prospect-of-big-fine-or-ban-if-twitter-fails-to-follow-new-legislation accessed on 08.12.2022.
[47] Mahdi Mirhoseini, Spencer Early, Nour El Shamy, Khaled Hassanein, Actively open-minded thinking is key to combating fake news: a multimethod study, Inf. Manage. 60 (3) (2023) 103761, https://doi.org/10.1016/j.im.2023.103761.
[48] a Patricia L. Moravec, Antino; Kim, Alan R. Dennis, Flagging fake news: system 1 vs. System 2, in: Jan Pries-Heje, Sudh Ram, Michael Rosemann (Hg.) (Eds.), Proceedings of the 39th International Conference on Information Systems. 39th International Conference on Information Systems, San Francisco, USA, 2018.
[49] Patricia L. Moravec, Antino; Kim, Alan R. Dennis, Appealing to sense and sensibility: system 1 and system 2 interventions for fake news on social media, Inf. Syst. Res. 31 (3) (2020) 987–1006, https://doi.org/10.1287/isre.2020.0927.
[50] Patricia L. Moravec, Antino; Kim, Alan R. Dennis, Randall K. Minas, Do you really know if it's true? How asking users to rate stories affects belief in fake news on social media, Inf. Syst. Res. 33 (3) (2022) 887–907, https://doi.org/10.1287/isre.2021.1090.
[51] Patricia L. Moravec, Randall K. Minas, Alan R. Dennis, Fake news on social media: people believe what they want to believe when it makes no sense at all, SSRN Electron. J. (2018), https://doi.org/10.2139/ssrn.3269541.
[52] Paul Mozur, A genocide incited on facebook. With Posts From Myanmar's Military, The New York Times, 2018. Hg. vThe New York Times. Online available under, https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html. accessed on 13.10.2022.
[53] David Murungi, Purao Sandeep, David Yates, Beyond facts: a new spin on fake news in the age of social media, in: Proceedings of the Americas Conference on Information Systems 2018, 2018.
[54] Eni Mustafaraj, Panagiotis Takis Metaxas, The fake news spreading plague, Hg, in: Peter Fox, Deborah McGuinness, Lindsay Poirer, Paolo Boldi, Katharina Kinder-Kurlanda (Eds.), Proceedings of the 2017 ACM on Web Science Conference. ACM Web Science Conference, New York, USA, 2017, pp. 235–239.
[55] Ka Chung Ng, Jie Tang, Dongwon Lee, The effect of platform intervention policies on fake news dissemination and survival: an empirical examination, J. Manage. Inf. Syst. 38 (4) (2021) 898–930, https://doi.org/10.1080/07421222.2021.1990612.
[56] Brendan Nyhan, Jason Reifler, When corrections fail: the persistence of political misperceptions, Polit. Behav. 32 (2) (2010) 303–330, https://doi.org/10.1007/s11109-010-9112-2.
[57] Anjan Pal, Snehasish Banerjee, Internet users beware, you follow online health rumors (more than counter-rumors) irrespective of risk propensity and prior endorsement, Inf. Technol. People 34 (7) (2021) 1721–1739, https://doi.org/10.1108/ITP-02-2019-0097.
[58] Anjan Pal, Alton Y.K. Chua, Dion Hoe-Lian Goh, How do users respond to online rumor rebuttals? Comput. Human. Behav. 106 (2020) 106243 https://doi.org/10.1016/j.chb.2019.106243.
[59] Gordon Pennycook, Tyrone D. Cannon, David G. Rand, Prior exposure increases perceived accuracy of fake news, J. Exp. Psychol. General 147 (12) (2018) 1865–1880, https://doi.org/10.1037/xge0000465.
[60] Gordon Pennycook, David G. Rand, Lazy, not biased: susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning, Cognition 188 (2019) 39–50, https://doi.org/10.1016/j.cognition.2018.06.011.
[61] Nornadiah Mohd Razali, Bee Wah Yap, Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests, J. Stat. Model. Analytics 2 (1) (2011) 21–33.
[62] Julio C.S. Reis, Andre Correia, Fabricio Murai, Adriano Veloso, Fabricio Benevenuto, Supervised learning for fake news detection, IEEe Intell. Syst. 34 (2) (2019) 76–81, https://doi.org/10.1109/MIS.2019.2899143.
[63] Jon Roozenbeek, Sander van der Linden, The fake news game: actively inoculating against the risk of misinformation, J. Risk. Res. 22 (5) (2019) 570–580, https://doi.org/10.1080/13669877.2018.1443491.
[64] Björn Ross, Anna-Katharina; Jung, Jennifer Heisel, Stefan Stieglitz, Fake news on social media: the (in)effectiveness of warning messages, in: Proceedings of 39th International Conference on Information Systems. 39th International Conference on Information Systems, San Francisco, USA, 2018.
[65] Victoria L. Rubin, Yimin; Chen, Nadia K. Conroy, Deception detection for news: three types of fakes, Proc. Assoc. Inf. Sci. Technol. 52 (1) (2015) 1–4, https://doi.org/10.1002/pra2.2015.145052010083.
[66] Swapnajeet Sahoo, Susanta Kumar Padhy, Jigyansa Ipsita, Aseem Mehra, Sandeep Grover, Demystifying the myths about COVID-19 infection and its societal importance, Asian J. Psychiatr. 54 (2020) 102244, https://doi.org/10.1016/j.ajp.2020.102244.
[67] Zhan Michael Shi, Gene Moo Lee, Andrew B Whinston, Toward a better measure of business proximity: topic modeling for industry intelligence, Manage. Inf. Syst. Q. 40 (4) (2016) 1035–1056.
[68] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, Huan Liu, Fake news detection on social media, ACM SIGKDD Explor. Newslett. 19 (1) (2017) 22–36, https://doi.org/10.1145/3137597.3137600.
[69] Melanie K.T. Takarangi, Sophie Parker, Maryanne Garry, Modernising the misinformation effect: the development of a new stimulus set, Appl. Cogn. Psychol. 20 (5) (2006) 583–590, https://doi.org/10.1002/acp.1209.
[70] Yuko Tanaka, Hiromi Arai, Miwa Inuzuka, Yoichi Takahashi, Minao Kukita, Ryuta Iseki, Kentaro Inui, Beyond click to cognition: effective interventions for promoting examination of false beliefs in misinformation, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems 961, 2025, pp. 1–18, https://doi.org/10.1145/3706598.3714243.
[71] The European Parliament and The Council of The EU, Regulation (EU) 2022/2065 of The European Parliament and of The Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act). Digital Services Act, Off. J. Eur. Union (2022). Online, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R2065. accessed on 15.04.2024.
[72] Bas Verplanken, Wendy Wood, Interventions to break and create consumer habits, J. Public Policy Market. 25 (1) (2006) 90–103, https://doi.org/10.1509/jppm.25.1.90.
[73] Nathan Walter, Riva Tukachinsky, A meta-analytic examination of the continued influence of misinformation in the face of correction: how powerful is it, why does it happen, and how to stop it? Commun. Res. 47 (2) (2020) 155–177, https://doi.org/10.1177/0093650219854600.
[74] Dandan Wang, Yadong Zhou, Yuxing Qian, Yunmei Liu, The echo chamber effect of rumor rebuttal behavior of users in the early stage of COVID-19 epidemic in China, Comput. Human. Behav. 128 (2022) 107088, https://doi.org/10.1016/j.chb.2021.107088.

[75] Christoph Weinert, Christian Maier, Sven Laumer, Tim Weitzel, Repeated IT interruption: habituation and sensitization of user responses, J. Manage. Inf. Syst. 39 (1) (2022) 187–217, https://doi.org/10.1080/07421222.2021.2023411.
[76] Ladd Wheeler, Harry T. Reis, Self-recording of everyday life events: origins, types, and uses, J. Pers. 59 (3) (1991) 339–354, https://doi.org/10.1111/j.1467-6494.1991.tb00252.x.
[77] Michael J.A. Wohl, Nyla R. Branscombe, Remembering historical victimization: collective guilt for current ingroup transgressions, J. Pers. Soc. Psychol. 94 (6) (2008) 988–1006, https://doi.org/10.1037/0022-3514.94.6.988.
[78] Helen Wyler, Margit E. Oswald, Why misinformation is reported: evidence from a warning and a source-monitoring task, Memory. 24 (10) (2016) 1419–1434, https://doi.org/10.1080/09658211.2015.1117641.
[79] Xia Zeng, Amani S. Abumansour, Arkaitz Zubiaga, Automated fact-checking: a survey, Lang. Linguist. Compass. 15 (10) (2021), https://doi.org/10.1111/lnc3.12438.
[80] Homero Gil de Zúñiga, Manuel Goyanes, Chris Skurka, Understanding fake news corrective action: a mixed-method approach, Int. J. Commun. 17 (2023) 3428–3450.

Christian Olenberger received his M. S. with honours in the elite graduate program Finance & Information Management at the Technical University of Munich, the University of Augsburg, and the University of Bayreuth, Germany. He received his doctorate degree in industrial engineering and management from the University of Augsburg. The focus of his work is on the design of digital technologies and assistance systems in the work and private lives of individuals.

Manfred Schoch is a Professor of Information Systems at Esslingen University of Applied Sciences. He is also a member of the FIM Research Center and the Business & Information Systems Engineering (BISE) branch of Fraunhofer FIT. He received his doctoral degree in economics from the University of Hohenheim. His research focuses on analyzing the use of digital technologies and media and deriving implications for their design to support human-centered digitalization.

Lena Utz received her M. S. with honours in the elite graduate program Finance & Information Management at the Technical University of Munich, the University of Augsburg, and the University of Bayreuth, Germany. She received her doctorate degree in economics from the University of Hohenheim. Her research focuses on human information behaviour.