Timing Matters: Designing Effective Corrections for Short-Form Video Misinformation

Suwani Gunasekara
School of Computing and Information Systems, University of Melbourne
Melbourne, VIC, Australia
gunasekara@student.unimelb.edu.au

Cherie Sew
School of Computing and Information Systems, University of Melbourne
Melbourne, VIC, Australia
csew@student.unimelb.edu.au

Saumya Pareek
School of Computing and Information Systems, University of Melbourne
Melbourne, VIC, Australia
saumya.pareek@student.unimelb.edu.au

Ryan M. Kelly
School of Computing Technologies, RMIT University
Melbourne, VIC, Australia
ryan.kelly@rmit.edu.au

Vassilis Kostakos
School of Computing and Information Systems, University of Melbourne
Melbourne, VIC, Australia
vassilis.kostakos@unimelb.edu.au

Jorge Goncalves
School of Computing and Information Systems, University of Melbourne
Melbourne, VIC, Australia
jorge.goncalves@unimelb.edu.au

Abstract

Short-form video platforms have become major channels for misinformation, with their rich multimodal features making false claims highly believable. HCI research shows that providing corrections in the same modality as the misinformation can be an effective solution. However, since corrections and misinformation convey contradicting information, the order in which one is exposed to them can impact what one believes. We conducted a between-subjects mixed-methods experiment where participants (N=120) rated the credibility of misinformation statements before and after viewing misinformation videos paired with correction videos. Corrections were shown either before, during, or after misinformation. Across all three timings, corrections reduced belief in misinformation, but post-exposure corrections proved most effective and mid-exposure corrections least effective. These findings suggest that correction mechanisms should appear after misinformation exposure, while avoiding mid-exposure interruptions that reduce impact. We outline design recommendations for integrating correction videos into short-form video platforms to improve resilience against misinformation.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**.

Keywords

Misinformation, Short-Form Videos, Timing, Corrections, Debunking, Social Media

ACM Reference Format:
Suwani Gunasekara, Cherie Sew, Saumya Pareek, Ryan M. Kelly, Vassilis Kostakos, and Jorge Goncalves. 2026. Timing Matters: Designing Effective Corrections for Short-Form Video Misinformation. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3772318.3790560

CHI '26, Barcelona, Spain

1 Introduction

The ability to convey messages through multimodal communication, creating a realistic and immersive experience in the viewer's mind, has made videos more persuasive than other modalities [113]. Historically, online video content predominantly existed in the form of long-format videos (e.g., YouTube). However, short-form videos have become increasingly popular due to platforms such as TikTok, YouTube Shorts, Instagram Reels, and Facebook Reels, making short videos one of the most widely consumed content formats online [72]. Due to the widespread reach and rapid consumption of short-form videos, these platforms have also increasingly become breeding grounds for misinformation [83].

Although platforms have a responsibility to ensure that users access and engage with credible information, their business models often prioritise user retention and engagement. This is achieved largely through mechanisms of virality and algorithmic amplification, which favour sensational content over accurate and reliable information [104]. As a result, platforms frequently neglect their responsibility to safeguard information credibility [5, 50, 121]. Instead of implementing direct and robust approaches, they tend to rely on softer moderation strategies, such as warning labels or links to third-party fact-checking sources [55, 120, 126]. For instance, TikTok uses banner warnings on suspect content to discourage users from sharing videos that may contain misinformation. When users encounter such videos, a banner appears near the top of the screen if the content has been reviewed by TikTok's fact-checking partners and could not be validated [46]. Similarly, YouTube has fact-checking information panels that display third-party verified articles above search results for relevant queries, providing additional context so users can make more informed decisions about the claims presented in news content [127].

While soft moderation techniques may work in more static or less dynamic environments, they are largely ineffective on fast-paced social media platforms [62]. Subtle nudges toward correct information often go unnoticed [38, 105], and even when users do see them, they require an additional step, such as clicking through to an external site to access accurate information. This disrupts the seamless experience that these platforms are designed to sustain, meaning most users simply ignore the correction [38]. More critically, misinformation on video-based platforms is consumed in a rich, multimodal format that is highly persuasive. Textual labels or external links cannot match this persuasive force. To truly counter misinformation in these rich, dynamically changing environments, corrections must be delivered in the same modality as the misinformation itself [37].

Video-based corrections, commonly referred to as *debunking videos*, operate by presenting factual information in video format to the original misleading content [7]. These videos are designed to directly counter false claims, often using the same visual and narrative techniques that make misinformation persuasive on video-sharing platforms. These correction videos are typically produced by experts or individuals with relevant expertise who aim to directly address the misinformation being circulated. A notable example is during the COVID-19 pandemic: as medical misinformation spread widely across social media, many healthcare professionals created and shared debunking videos to counter false claims and provide accurate, evidence-based information [104].

Currently, debunking videos may be encountered at unpredictable points in a user's online journey due to the way they are distributed. As a result, some users are exposed to a correction video before encountering the misinformation, while others only see it afterwards, if at all. The timing of corrections is critical as it shapes how people process and evaluate information. A correction seen before exposure to misinformation may act as a "prebunking" or inoculation, preparing viewers to resist false claims [12, 100]. A correction delivered after exposure could act as a form of feedback, enhancing long-term learning and memory [44]. If a correction is presented during the exposure to misinformation, it can interrupt the processing of the false content while simultaneously offering an immediate alternative explanation. This could also support knowledge revision [12], as people are more likely to update their understanding when the misinformation and the correction are coactive [61]. The simultaneous presentation allows users for direct comparison, making it easier to replace the false claim with the accurate one. Taken together, these perspectives raise a central question: ***when is the most effective time to present correction videos - before, during, or after exposure to misinformation?***

Despite extensive work on misinformation corrections, most research on the timing of corrections has been conducted in text-based contexts, with mixed results reported [12, 24, 97, 98]. While some studies find that corrections are most effective when presented after misinformation exposure [12, 24], others report that timing has minimal or no influence on how well corrections work [97, 98]. At present, there is no clear understanding of when corrections should be presented in short video environments. While insights from text-based studies are valuable, it is unclear if they can be directly applied to dynamic, rich-media short video platforms like TikTok. Unlike text, short-form videos deliver information rapidly across visuals, audio, and narrative, heightening attentional capture and emotional engagement [99]. This multimodal presentation makes video content more immersive and memorable [45], and also fundamentally changes how information is processed [19]. Studies of short-form video platforms (e.g., TikTok) further show that design features such as autoplay, continuous feeds, and tightly edited clips encourage rapid, low-effort consumption and emotional engagement, which differs from the more deliberative, self-paced nature of reading [84]. As a result, the optimal timing of corrections in video contexts *may* differ from text-based settings, where cognitive demands and attention patterns are not the same.

Thus, in this paper, we examine how the timing of a correction influences the perceived credibility of misinformation in short-form videos. To achieve this, we conducted a between-subjects experiment with 120 participants. Participants viewed and rated the perceived credibility of misinformation statements both before and after exposure to misinformation videos. These videos were presented with an accompanying correction video, introduced at one of three timing conditions: prior to the misinformation video (T_{start}), during the misinformation video (T_{mid}), or after the misinformation video (T_{end}). By analysing the credibility ratings, we examined the likelihood that participants would reduce their trust in misinformation statements across the different timing conditions, in order to identify which timing was most effective. We concluded our experiment by asking participants open ended questions to further understand their preferences for the correction placement and how it influenced their interpretation of the misinformation statements. We found that providing a correction video at any of the three timings was effective in reducing belief in misinformation. However, corrections were most effective when presented after the misinformation video. To further contextualise these findings, we explore participants' qualitative reflections on the correction timings, offering insights into why certain placements were perceived as more effective than others. We conclude by discussing how our findings can be used to mitigate the spread of misinformation in short-form video platforms.

Our work contributes to misinformation research in three key ways. First, by systematically testing correction timing in the context of short-form videos, we fill an unexplored research gap that has predominantly examined text-based corrections. Second, we empirically show that for short-form video misinformation, corrections are most effective when delivered *after* the misleading content rather than *during* it. Third, drawing on this finding, we outline how platforms can incorporate timing-sensitive design strategies to improve the delivery and effectiveness of video-based corrections.

2 Related Work

Misinformation is defined as information that is inaccurate according to expert consensus at the time of dissemination [123], and shared regardless of intent [4, 56]. Although the dissemination may lack malicious intent, misinformation spreads rapidly; particularly on social media platforms, amplifying its potential impact. The negative effects of misinformation are evident at both individual and societal levels [29], as it can lead to ill-informed decisions with serious economic, social, and health-related consequences [9, 69, 71]. In the following sections, we start by examining the literature on tackling misinformation on short-form video platforms. We then highlight the role of corrections as a key approach to reducing the influence of online misinformation.

2.1 Misinformation in Short-Form Video Platforms

In the past, misinformation was primarily disseminated in text format, largely due to the text-based nature of early social media platforms and the relative ease with which text-based misinformation could be created and shared. However, with the rapid advancement of technology and the evolution of social media, the ways in which people consume information have shifted significantly. Today, users increasingly rely on media formats such as images and videos to communicate, creating new dynamics in information exchange. According to Media Richness Theory [23], the effectiveness of information delivery is shaped by the richness of the communication medium, defined by its capacity to convey multiple cues, provide immediate feedback, and express personal context. Rich media such as images, audio, and videos have therefore become dominant modes of communication, rendering misinformation more engaging, persuasive, and difficult to detect [2, 65].

In videos, visual and auditory information are synchronised, making message processing more natural and closer to real-life experiences, which in turn enhances perceived truthfulness [124]. Moreover, videos employ various persuasive strategies such as engaging visuals, compelling audio, and authoritative narration that make the information appear more coherent and convincing [28]. Consequently, video-based misinformation tends to be shared more widely than text- or audio-based misinformation, driven by this heightened sense of perceived credibility [113].

Videos can be particularly effective because their multisensory nature supports long-term information retention, as explained by dual coding theory [19]. This theory posits that the brain processes information through two interconnected channels; verbal and visual. By engaging both simultaneously, videos strengthen memory recall. In fact, prior research has shown that learners who processed visual and linguistic information simultaneously retained nearly twice as much as those who focused only on auditory or linguistic input [22].

These same multimodal features also make misinformation detection more difficult. Assessing accuracy in rich-media content requires evaluating not just text, images, or audio individually but also the combined meaning they convey [2]. A benign statement paired with an unrelated but emotionally charged image can produce a misleading implication even if neither element is factually incorrect on its own. Although numerous detection models have been developed in the past decade [54, 106], most rely on a single modality, such as text [36, 47] or images [1, 51] and therefore miss crucial cross-modal cues. Ensemble approaches that merge independently trained models across modalities [39, 107] offer some improvement, but simply combining separate modalities often fails to capture how those elements work together to create misleading messages [2]. These limitations help to explain why multimodal misinformation is particularly challenging to detect, especially on dynamic audiovisual platforms like TikTok. A platform originally designed for entertainment content such as dance and music, now hosts a rapidly growing volume of misinformation [41].

Given the highly persuasive and potentially harmful nature of video-based misinformation, the HCI community has increasingly turned its attention to this issue. Research in this space has examined several key avenues, including how misinformation spreads on video platforms [15, 52, 88, 122], how users perceive and engage with such content [3], how they assess its credibility [43], and what types of interventions may be effective in mitigating its influence [38, 41]. Prior work shows that users often *encounter* misinformation incidentally rather than actively *seeking* it [43]. While platforms employ de-recommendation strategies to reduce exposure, which can decrease the spread of targeted content, such measures may inadvertently draw attention to related conspiratorial material [15]. Hassoun et al. [43] argue that the most effective approaches embed solutions and strategies for addressing misinformation directly within the social contexts of online interaction.

Currently, short-form video platforms, including TikTok and YouTube Shorts, primarily rely on soft moderation strategies. Instead of removing misinformation outright, these approaches allow content to remain on the platform while attaching informative labels or contextual notes to indicate the material may be misleading. Platforms often favour this approach because hard moderation, such as removing content or suspending accounts, can create perceptions of censorship, position the platforms as arbiters of truth [34], and disrupt their business models that depend on maximising user engagement and retention. However, existing soft moderation techniques face important limitations: they are often not prominently visible, making them easy to overlook [105], and when clicked, corrective links frequently redirect users to external sources. This additional step requires effort, making it less likely that users will actually engage with corrective information [38], thereby reducing the overall effectiveness of such interventions. Given these challenges, researchers have increasingly turned to direct correction strategies as one of the most promising ways to mitigate misinformation.

2.2 Correcting Online Misinformation

Prior research in HCI has examined a variety of strategies to address online misinformation. One of the most popular and effective strategies is the use of corrections, commonly referred to as *Debunking*. Debunking is defined as the mechanism of exposing the incorrectness of the misinformation claims [70]. Unlike interventions such as warning labels or fact-checking links, which primarily serve as nudges encouraging users to seek accurate information, debunking directly addresses misinformation by presenting corrective content. Rather than relying on users to pursue additional sources or reinterpret information on their own, debunking supplies the accurate explanation alongside the false claim, thereby reducing ambiguity and offering a clear alternative narrative, making debunking a powerful strategy against misinformation [104].

Previous work has explored different dimensions of corrections to better understand their mechanisms and the factors that contribute to their effectiveness. One concern has been whether repeating misinformation during a correction might backfire by increasing its familiarity. However, Ecker et al. [29] found no evidence of such a familiarity backfire effect, suggesting that it is both safe and often useful to restate misinformation before correcting it, as this ensures clarity about which claim is being refuted. At the same time, the effectiveness of fact-checking depends on the credibility of the source delivering it, with prior work showing that users are more receptive when the source is familiar or trustworthy [40, 75]. Interestingly, Bode et al. [10] demonstrated that even observing someone else being corrected can reduce belief in misinformation. The tone of corrections, however, appears less important: Martel et al. [76] found that whether corrections were delivered in a direct or hedged style, and whether they offered simple or detailed explanations, these stylistic differences did not significantly influence engagement or acceptance. Similarly, Kotz et al. [67] compared two structural approaches, "bottom-heavy" corrections (misinformation first, followed by correction) and the "truth sandwich" (accurate information, misinformation, then correction), and likewise reported no significant differences in effectiveness.

As correction and misinformation present conflicting information, the sequence in which they are delivered can significantly affect individuals' memory and comprehension [24]. Therefore, the timing of the corrective information may be a critical factor in the overall effectiveness of debunking efforts. However, the findings in this matter remain mixed. Along with examining the coherence between correction and misinformation, Dai et al. [24] investigated whether the timing of text correction; before versus after text misinformation affected its corrective impact. The results showed that corrections were more effective when presented after the misinformation, especially when paired with a coherence-enhancing message. These effects were not only immediate but also persisted for at least one week following the initial exposure. A similar finding was also reported by Brashier et al. [12] when the authors examined whether fact-checking labels placed before, during, or after exposure to misinformation articles were most effective. They also reported that the most effective timing is seeing it after the exposure to misinformation. However, Rich and Zaragoza [98] found no evidence that the timing of correction significantly influenced its effectiveness. In their study, corrections were delivered either minutes after initial exposure to misinformation or after a delay of two days. Regardless of the timing, the immediate effectiveness of the correction remained unchanged. In contrast, Craig and Vijaykumar [21] observed that when infographic-based corrections were followed shortly by re-exposure to the same misinformation, the initial corrective effects weakened, suggesting that misinformation can 'undo' the correction if it is encountered soon afterward. Nevertheless, Prike and Ecker [97] argue that the order of presentation may matter less, reporting that corrections are likely to be similarly effective whether delivered pre-emptively (i.e., prebunking) or after exposure, with timing having limited impact on overall effectiveness.

While these findings provide valuable insights, most prior research has focused primarily on text-based misinformation. Yet misinformation is no longer confined to textual formats; it has increasingly evolved into richer modalities such as video. Video-based misinformation poses distinct challenges because it is often more engaging, persuasive, and difficult to scrutinize compared to text, making it potentially more harmful. For example, Wittenberg et al. [124] found that individuals were more likely to believe a political event occurred when the information was presented in video rather than text. Similarly, Sundar et al. [113] found that video-based news is often perceived as more credible and is more likely to be shared compared to textual formats. This makes misinformation videos particularly dangerous, as they can more easily create false memories and lead individuals to make misinformed decisions [14]. At the same time, prior research has shown that video corrections are more effective in countering video misinformation [7, 37] and are also more impactful than lengthy fact-checking articles in correcting misinformation [125].

Existing studies highlight the advantages of modality-matched corrections, yet an important question remains under-explored: does the timing of a correction within a video influence its effectiveness? Although prior work on text-based misinformation has examined correction timing, its mixed and contradictory findings offer little guidance for video-based misinformation. Moreover, the multimodal, rapid, and attention-fragmented nature of short-form videos warrants additional investigation into the appropriate timing of corrections. Understanding when a video correction should appear is crucial, as timing may affect not only whether viewers notice the correction but also how they process and integrate it into their beliefs. Despite the growing prevalence of video-based misinformation, no prior research has systematically investigated the temporal dynamics of video corrections. This study addresses this gap by examining how the timing of corrections influences their effectiveness in reducing the believability of video misinformation.

3 Method

To explore whether the timing of short-form video corrections influences perceived credibility change, we conducted a between-subjects survey-based experiment employing a mixed-methods design. This approach enabled us to capture quantitative measures of perceived credibility change and qualitative insights into how the timing of corrections affected participants' beliefs and their preferences regarding the intervention. All procedures received approval from our university's Human Research Ethics Committee.

3.1 Stimuli

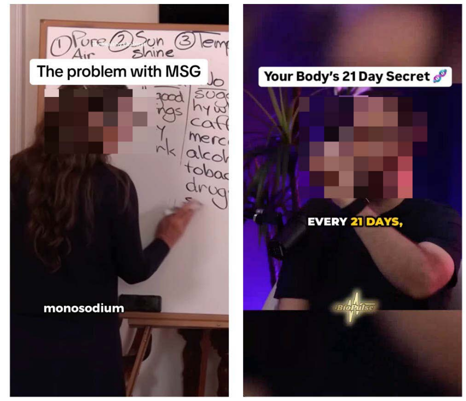
Our study involved participants rating the credibility of misinformation statements before and after seeing misinformation and correction videos. To systematically develop our stimuli, we first formulated a set of criteria to guide the selection process. These criteria were intended to ensure that the misinformation statements and the accompanying videos were both realistic and believable, while also minimising potential confounds that could bias participants' responses. The following subsections detail the criteria applied in the selection of misinformation statements and the corresponding videos.

3.1.1 Misinformation Statements. The misinformation statements used in the study were selected according to several criteria to ensure their suitability. First, the statements could not be overly politically charged or ideologically polarising, as exposure to politically sensitive topics can activate participants' pre-existing attitudes and thereby confound the results [116]. To operationalise this criterion, we excluded statements that referenced contemporary political actors, public policy debates, or issues empirically shown to map onto partisan or identity-based divisions such as climate change [6, 109] or vaccination [53], which are known to elicit strong affective and ideological responses. In line with evidence that identity-based cues can influence judgement even in contexts that are seemingly non-political [101], two researchers independently screened an initial pool of candidate statements and removed any items containing explicit or implicit ideological cues. This ensured that the remaining statements would be evaluated primarily on their perceived plausibility rather than on partisan alignment. Second, we excluded widely debunked misconceptions, such as the Flat Earth theory, since these are generally recognised as false and would not provide meaningful variation in believability. Third, statements that were non-harmful or trivial, for example, misconceptions that do not pose risks to health, safety, education or public well-being, were also avoided, as the study focused specifically on misinformation with potential real-world consequences. Finally, the selected statements needed to be plausible enough that participants could reasonably engage with them, but not so obviously false that they would be dismissed without consideration. Below are the statements that we used in the study, accompanied by references to the accurate information that was shown to participants.

- 5G is harmful to human health [27]
- MSG is an unsafe flavor enhancer that poses risks to overall health [82]
- Microwaves cause cancer and other serious health issues due to radiation [66]
- People can be divided as left and right brain people [85]
- Individuals learn better when they receive information in their preferred learning style [111]
- It takes 21 days to form a habit [86]

3.1.2 Videos. The misinformation and correction videos were sourced from content already circulating on TikTok, YouTube Shorts, and Instagram Reels. To guide selection, we developed a set of criteria aimed at minimising bias and ensuring consistency across all stimuli. To avoid potential influences stemming from familiarity or perceived authority, presenters in the correction videos were chosen such that they were not widely recognised, famous, or easily identifiable with a specific profession. This helped ensure that participants' responses were based on the content of the message rather than the perceived credibility of the speaker. In terms of length, videos were constrained to a duration between 30 and 90 seconds, sufficient to convey a clear explanation while maintaining participant engagement [74]. This video length also is typical for content found on short video platforms [25]. All videos were screened to exclude offensive language or inappropriate material, ensuring that content was suitable for an experimental setting. Importantly, narration within the videos needed to provide a logical and coherent explanation that either supported or refuted the specific misinformation statement, allowing participants to process corrective information effectively. Furthermore, we limited the selection to videos with low view and like counts to ensure a minimal chance of prior exposure among participants. Finally, videos that relied solely on background music without narration were excluded, as the absence of verbal explanation would limit the capacity to deliver meaningful corrections. Screenshots of some of the videos used in the study can be seen in Figure 1. A summary of the videos used in the study can be seen in Appendix A. Our stimuli are also available to the research community[1]. To preserve presenter anonymity, their faces are blurred in the shared materials, although they were visible to participants during the actual study.

3.1.3 Overlay Design. We designed overlays to create a smooth transition between the misinformation and corrective content and also back to misinformation content. Rather than abruptly interrupting the misinformation video, the overlay served as an intermediate



(a) Misinformation Videos

(b) Correction Videos

Figure 1: Example screenshots of misinformation and corresponding correction videos used in the study.

step that prompted participants to engage with the correction at the designated time point (T_{start}, T_{mid} or T_{end}). Both overlays contained a single button that allowed participants to navigate to the correction video (Figure 2a) and then return to the misinformation video (Figure 2b), as illustrated in Figure 2. To support ecological validity, the overlays were designed with a translucent effect similar to those commonly used in social media platforms.

The design of the overlay shown before the correction video incorporated an icon similar to that used in Community Notes on the platform X (formerly Twitter), in order to convey the impression that the prompt originated from users rather than from the platform itself. To reinforce this framing, the accompanying text explicitly stated that the video had been flagged by other users, rather than by fact-checkers or platform moderators. This design choice was informed by prior research showing that users tend to distrust platform-led fact-checking interventions, but are more receptive to peer-driven cues [110]. Furthermore, we used the term 'debunking video' instead of 'correction video', as using the word correction would have implied that the initial statement was definitively false, potentially biasing participants' responses. In contrast, the term debunking framed the video as a counter-argument rather than a

[1] https://osf.io/hdtac/overview?view_only=08ecedd8f80a4c469cd1ac337bb02ce9

predetermined judgment, allowing participants to engage with the content more naturally.

3.2 Experimental Design

While prior work has shown that video corrections can reduce belief in misinformation [7, 37], the question of when these corrections should appear within a video remains largely unexplored. To address this gap, we designed our experiment to examine whether participants would be most likely to reduce perceived credibility in misinformation when a correction was shown before, during, or after exposure.

Since prior work already demonstrates that corrections are effective, we held their presence constant and manipulated only their temporal placement, focusing on the relative effects of correction timing. We therefore employed a between-subjects design with three conditions: correction at the start, middle, or end of the video. We first measured both pre- and post-exposure perceived credibility. To capture pre-exposure credibility judgments, participants were first presented with the six misinformation statements and asked to rate their level of agreement with each. Consistent with prior work on misinformation correction [63, 64, 112], these ratings were collected using a 7-point Likert scale *(1 = Strongly Disagree, 7 = Strongly Agree)*.

When presenting corrections, we ensured that participants were shown videos addressing the statements they found most believable, as correcting claims they had already rejected would provide little insight into credibility shift. To achieve this, we generated an internal ranking for each participant based on their perceived credibility scores (e.g., a statement rated 6 ranked higher than one rated 5). The two highest-ranked statements were then paired with corresponding correction videos, though the ranking process itself remained invisible to participants. To proceed to the next section of the survey, participants were required to have rated at least two statements above 1 on the scale; those failing to meet this criterion were excluded from the study.

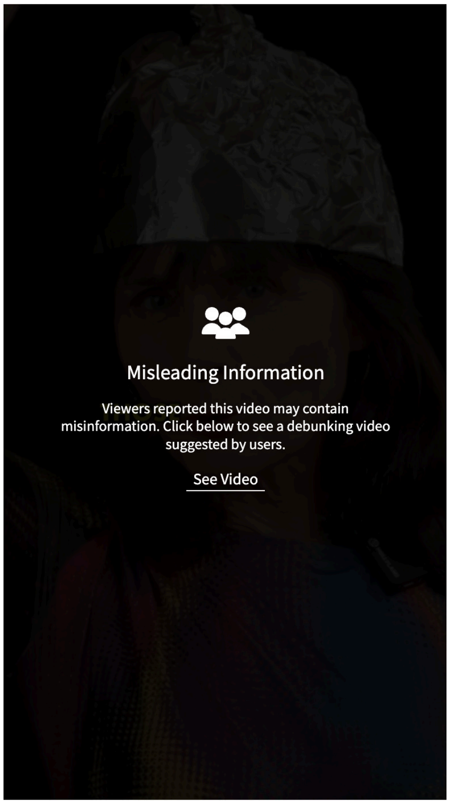
The next section of the study consisted of eight short videos: six misinformation videos and two correction videos, each ranging from 30 to 90 seconds in length. Every misinformation video corresponded to one of the misinformation statements shown in the first section, while the two correction videos addressed the statements that participants had ranked highest in believability. These two correction videos were paired with their corresponding misinformation videos, based on the *timing* condition; T_{start}, T_{mid} or T_{end}. To minimise potential order effects, the videos were presented in a randomised sequence in our experiment. The two correction videos always appeared at fixed positions (2 and 5) within the sequence, but the assignment of which high-belief statement appeared in each position was randomised across participants.

As seen in Figure 3, in the T_{start} condition, the correction video was shown before the participant viewed the corresponding misinformation video. In the T_{mid} condition, the correction video was played after approximately 50% of the misinformation video had been watched. In the T_{end} condition, the correction video appeared immediately after the misinformation video had finished. In all three scenarios, an overlay mechanism was used to ease participants into the transition, rather than abruptly placing the correction video on top of the misinformation video. This design ensured that participants engaged with the corrective content at the intended time point without disrupting the natural flow of video consumption. After watching the correction video in all three timing conditions, participants were prompted via an overlay to return to the misinformation video. They were then required to finish any remaining portion of the misinformation video before proceeding to the next question. The detailed design of the overlay is provided in Section 3.1.3. To maintain a consistent level of friction across conditions, participants in all three time conditions encountered both overlays and always concluded by returning to the misinformation video (see Figure 3). Finally, to ensure full exposure to the stimuli, navigation buttons were configured so that participants could proceed to the next question only once the video(s) had fully finished playing.

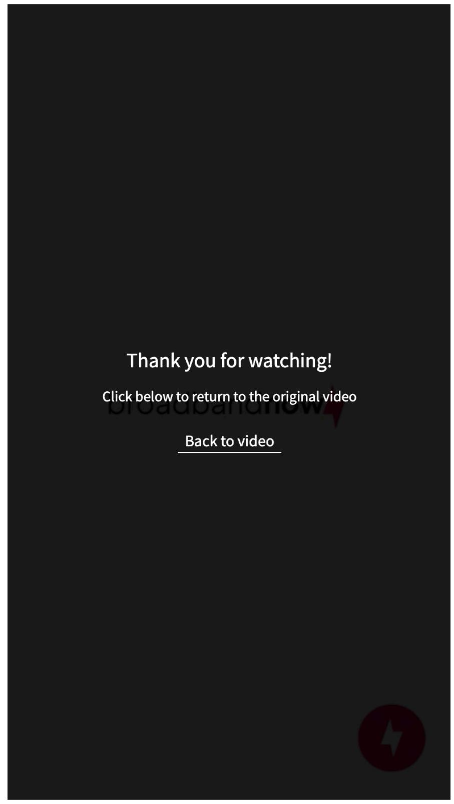
The final section of the study included open-ended questions designed to explore participants' perceptions of the timing of the correction videos. Specifically, participants were asked to reflect on the timing at which they viewed the correction, as well as to consider the other two possible timings and how their perceptions might have differed had the correction been presented at those points instead.

For misinformation intervention research, it is essential to account for individual differences, such as deliberative thinking style, which can influence the perception of misinformation and potentially confound results [116]. Following past literature on misinformation [37, 76, 92, 94, 117], to isolate the true impact of the interventions, we controlled for these differences by employing both the Bullshit Receptivity Scale (BSR) [89] and the Actively Open-Minded Thinking scale (AOT) [108].

The BSR, developed by Pennycook et al. [89], consists of 10 pseudo-profound statements incorporating abstract buzzwords. Participants were instructed that "profound" refers to "of deep meaning; of great and broadly inclusive significance" and were asked to rate the perceived profoundness of each statement on a five-point Likert scale (1 = Not at all profound to 5 = Very profound). We hypothesise that participants with higher BSR scores will show a greater tendency to believe misinformation, as a higher receptivity to pseudo-profound statements may reflect a reduced tendency to critically evaluate content. The AOT scale, based on the framework by Stanovich and Toplak [108], measured the tendency to be open towards opinions different from one's own. Participants rated 13 statements on a six-point Likert scale (1 = Strongly disagree to 6 = Strongly agree). We hypothesise that participants with higher AOT scores will be more receptive to corrections than those with lower scores. One Instructional Manipulation Check (IMC) was included in the online survey to ensure participant attentiveness [87]. Participants were instructed to select 'Somewhat profound' in one of the questions while completing the BSR questionnaire. No IMC was embedded during the main part of the study (i.e., while participants watched misinformation and debunking videos and provided their ratings) to avoid disrupting the natural flow of the viewing experience. The IMC item was removed prior to computing participants' mean BSR scores for analysis. The study procedure is summarised in Figure 4.



(a) Overlay shown before the correction video



(b) Overlay shown after the correction video

Figure 2: Overlay designs used in the study. Figure 2a shows the overlay presented before the correction video, prompting participants to view the debunking video. Figure 2b shows the overlay displayed after the correction video, prompting participants to return to the original misinformation video. Both overlays were shown in all three timing conditions.

3.3 Participants and Data Collection

The study was administered through Prolific[2]. Participants were required to be fluent in English, to have a Prolific approval rating above 98% and to be frequent users of short video platforms, e.g. TikTok, YouTube Shorts or Instagram. Furthermore, participants were limited to those residing in the United States, as TikTok reportedly holds the largest market share in this region [118]. A minimum of 116 participants was recommended via a G*Power Analysis [31] considering an $\alpha = 0.05$, and a power of 0.9. Hence, we recruited a total of 120 participants (60 men and 60 women; M = 42.93 years, SD = 12.26, range = 21–83) to cover the three conditions in a balanced manner. Participants spent a median time of 15 minutes on the survey and received US$3 for their participation.

We used the Qualtrics platform[3], which enabled controlled stimulus presentation and efficient data collection from a relatively large sample, an approach commonly adopted in misinformation research [42]. After providing informed consent and completing demographic questions, participants were randomly assigned to one of three conditions. As the study employed a between-subjects design, 40 participants were assigned to each condition: T_{start}, T_{mid}, and T_{end}. In total, eight participants (T_{start}: 2, T_{mid}: 4, T_{end}: 2) did not progress to the video-watching section of the survey because they failed to meet the requirement of rating at least two statements above 1 on the pre credibility rating likert scale. Consequently, 8 additional participants were recruited to replace them.

[2]https://www.prolific.com/
[3]https://www.qualtrics.com/

4 Results

Each participant was shown a total of eight videos: four misinformation videos presented without any accompanying debunking content, and two misinformation videos that were paired with corresponding correction videos. We first present quantitative findings on whether showing a correction video reduces perceived credibility and whether the timing of the correction influences its effectiveness. We then complement these findings with a qualitative analysis of participants' reflections, offering insight into why certain correction timings were perceived as more effective.

4.1 Quantitative Analysis

We first examined the overall effect of presenting a correction video to participants in order to assess whether the intervention

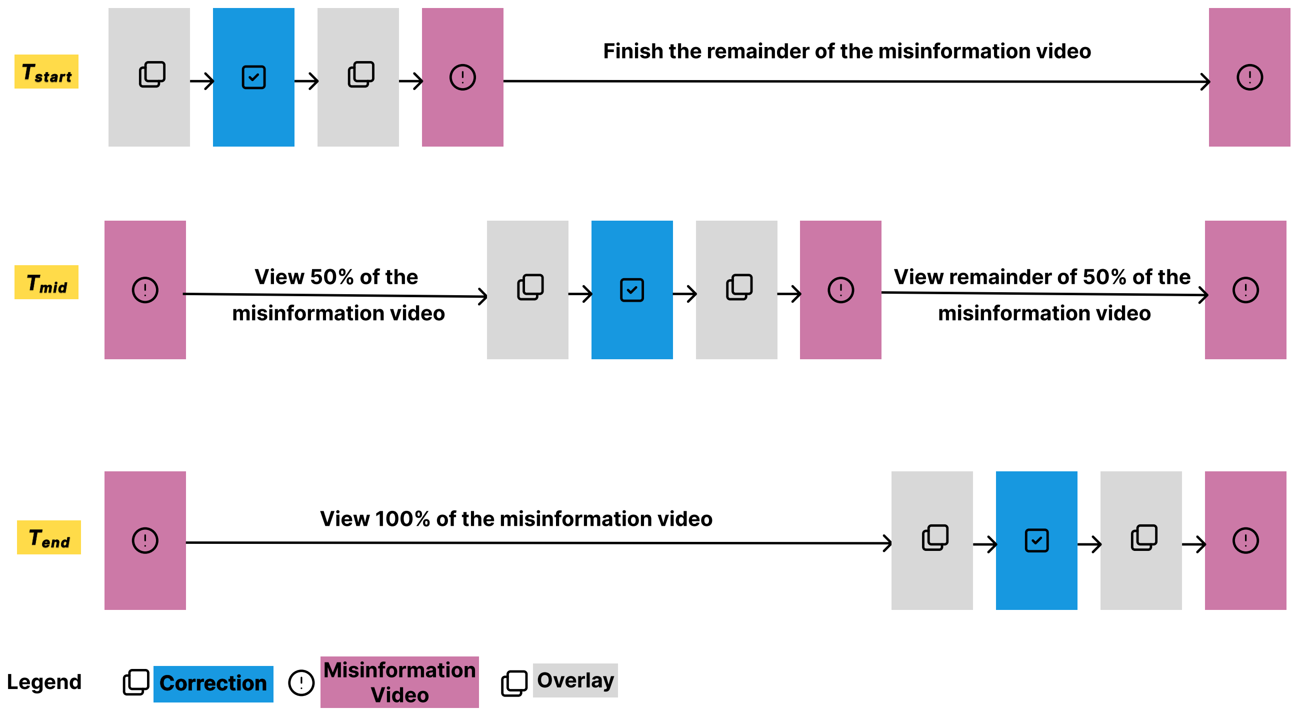


Figure 3: The placement of corrections and misinformation videos across the three conditions in our study. To keep the level of friction consistent across conditions, all participants encountered both overlays regardless of timing, always ending by returning to the misinformation video. In the relevant time condition, participants first saw an overlay indicating that the video had been reported as containing misinformation. They were then prompted to click the "See Video" button to view the debunking content. Upon clicking, the correction video was shown, and after it ended, another overlay appeared to allow participants to navigate back to the misinformation video.

generally reduced belief in misinformation. We then analysed the between-conditions effect of correction video timing, focusing on whether presenting the correction at the start, middle, or end of the misinformation video altered its effectiveness.

4.1.1 Effect of showing the correction video. We analysed the number of instances in which participants reported a reduction in belief following the correction intervention. A belief was coded as **Reduced** if the participant's perceived credibility rating after watching the video was lower than their rating before the video (i.e., post rating < pre rating).

As shown in Figure 5, approximately 70% of responses had a reduction in believability of the misinformation statement when the corresponding video was paired with a correction video, compared to only about 13% when no correction video was shown. This finding indicates that participants were considerably more likely to reduce their perceived credibility in a misconception when a debunking video was present.

To formally test this relationship, we fitted a Generalized Linear Mixed Model (GLMM) with a binomial distribution using the lme4 package in R. The model included **Correction** as a fixed effect predictor, and random intercepts for both participant ID and statement index to account for repeated measures across participants and variation across different statements. The results indicated a significant positive effect of the debunking intervention on perceived credibility reduction (β = 3.82, SE = 0.33, OR = 45.6, z = 11.55, p < .001). This suggests that the odds of perceived credibility reduction were approximately 45.6 times higher when a correction video was present.

4.1.2 Effect of timing. To investigate whether the timing of the correction video had a significant effect on the reduction in perceived believability, we restricted our analysis to instances in which participants were shown a correction video. This resulted in a total of 240 observations (2 instances per participant × 120 participants). After removing outliers (1% of the data), we fitted another GLMM.

Model Building: The outcome variable **Reduced** was defined as a binary indicator, representing whether participants rated the believability of the statement lower after viewing the debunking video. Prior research has shown that age [13, 60] and frequency of social media use [26, 79] influence susceptibility to misinformation. Therefore, we included **Age** of the participant and frequency of video-sharing platform **Usage** as fixed predictors in our model. We

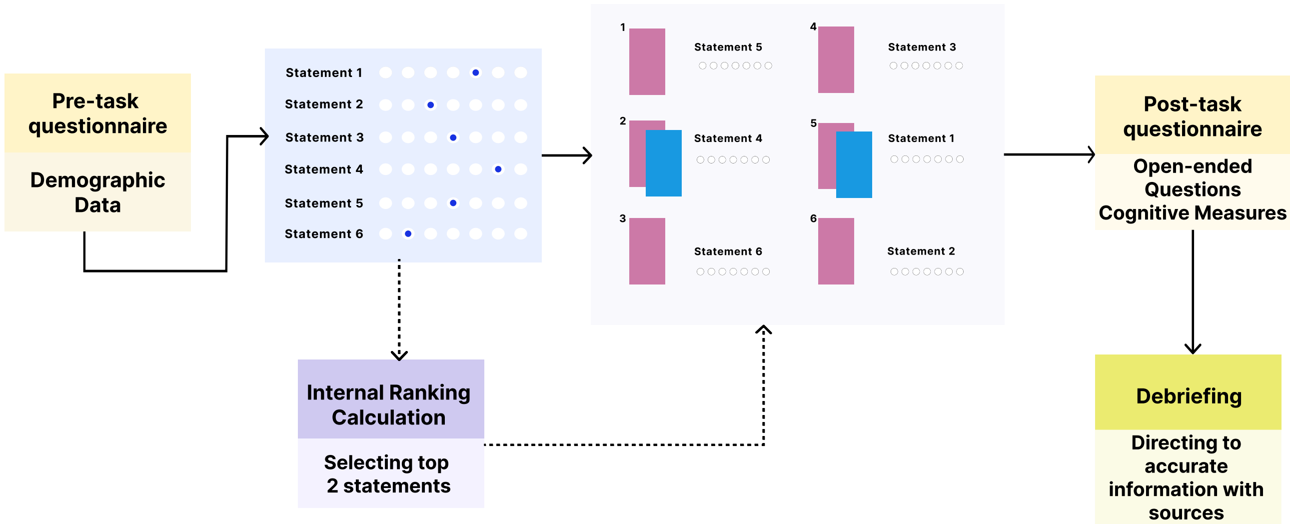


Figure 4: Flow diagram of the study procedure. After completing the demographic details, participants rated the believability of six misinformation statements using a 7-point likert scale. Following this, they were shown a series of short videos corresponding to the same misinformation statements, with the two highest believability statement videos being paired with a correction video. After each video, participants again rated the believability of the related statement using the same 7-point likert scale. After completing all video segments, participants responded to open-ended questions designed to collect qualitative data about their perceptions of the timing of the debunking videos. They also completed two additional measures: the Bullshit Receptivity Scale [89] and the Actively Open-Minded Thinking Scale questionnaires [108]. At the conclusion of the survey, participants were thanked for their time and provided with fact-checked links to the stimuli for debriefing purposes.

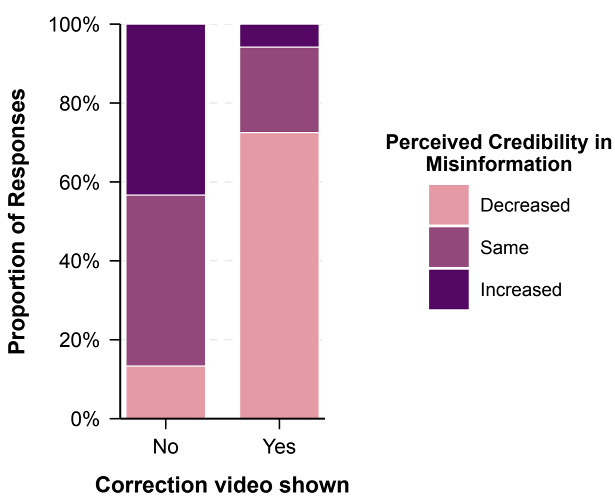


Figure 5: Effect of correction intervention in change in belief ratings.

also incorporated **BSR** and **AOT** as fixed predictors, and participant ID and video ID as random effects in the model to account for individual differences and potential variability associated with specific statements/videos. Diagnostic checks indicated that there was no multicollinearity among the predictor variables. Furthermore, the random effects were approximately normally distributed and independent of the response variable, supporting the appropriateness of the model assumptions.

The results of the GLMM analysis are presented in Table 1 and illustrated in Figure 6. As shown in Figure 6(a), corrections presented at the end of the misinformation video were most effective in belief reduction. Compared to the T_{end} condition, participants at T_{mid} condition were significantly less likely to reduce perceived credibility (*β = -1.00, SE = 0.47, p < 0.05*), whereas T_{start} (*β = -0.40, SE = 0.48, p < 0.361*) condition did not differ significantly from T_{end}. Finally, we observed a significant main effect of Age (*β = 0.454, SE = 0.226, p = 0.044*). Older participants were less likely to reduce perceived credibility in response to correction videos, as seen in Figure 6(b).

4.2 Qualitative Analysis

To further explore our main research question, we qualitatively analysed participants' open-ended questionnaire responses. The analysis aimed to understand the participants' perceptions about the effectiveness of particular correction timings, and whether the

Table 1: Fixed-effect estimates from the generalized linear mixed model. Statistically significant main effects (p < .05) are shown in bold. The sign of the estimate (+/-) indicates the direction of the association: positive estimates correspond to a higher probability of reducing believability, whereas negative estimates correspond to a lower probability.

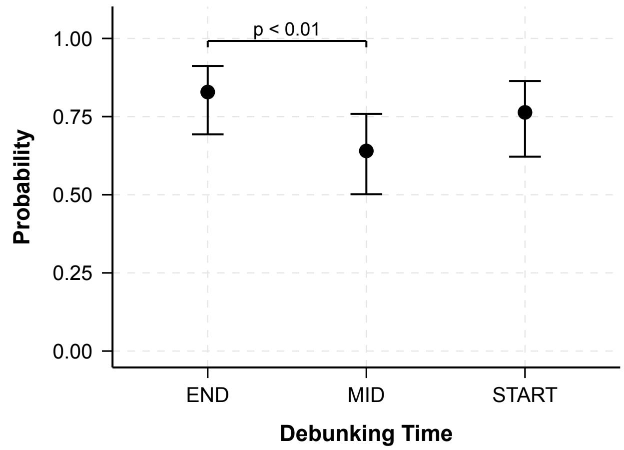
Predictor	β (Std. Error)	p value
$(Intercept)$	1.496 (0.399)	**<0.001**
T_{Mid}	-1.001 (0.471)	**0.030**
T_{Start}	-0.407 (0.480)	0.361
AOT_{Mean}	0.415 (0.217)	0.056
BSR_{Mean}	-0.339 (0.208)	0.106
$Usage$	-0.021 (0.219)	0.922
Age	-0.454 (0.226)	**0.044**

participants would have changed their perceptions about the misinformation if the correction video had been shown at a different time. This complements our quantitative analysis by providing an understanding of why the video endpoint timing was most effective, while considering the plausible benefits of other placements.

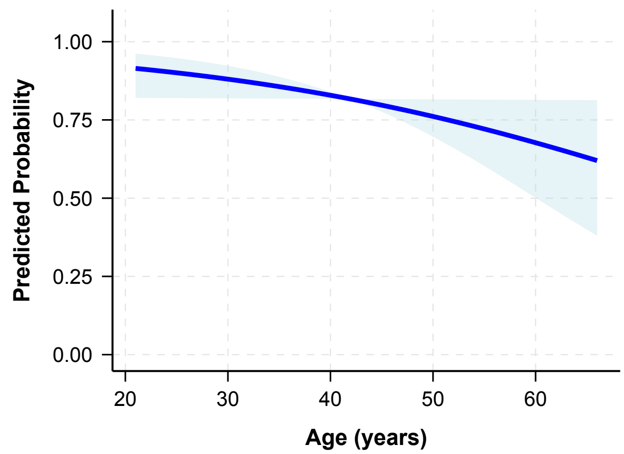
We analysed the qualitative responses using a general inductive approach [119]. The first author began by reading the responses and developed a preliminary set of categories representing participants' opinions about the timing of the correction video. These categories were iteratively refined in collaboration with another member of the research team. Subsequently, both coders independently applied the final set of codes back to the participant responses. We assessed inter-rater reliability using Cohen's Kappa, yielding a value of 0.81, indicating strong agreement between coders [78]. Disagreements were resolved through discussion.

4.2.1 T_{end}: Out of the 40 participants who were in this condition, 34 indicated that their perceptions would not have changed had the correction been shown at a different time. The main benefit of end placement was that it was seen as less disruptive and more natural, allowing participants to watch both clips in their entirety. For example, one participant noted, *"... at the end is the perfect place because it allows you to watch both full video clips."* ($P90_{end}$). In contrast, 5 participants felt that their perceptions would have changed if the correction had been shown at a different time. One suggested that presenting the correction at the start would have been more effective, reasoning that initial exposure shapes belief: *"... we don't have anything as reference so people's brains are wired to believe that whatever they hear first is perceived to be the truth and tend to hang on it."* ($P95_{end}$). The remaining participant was uncertain of how they might have reacted to a change in video placement.

4.2.2 T_{mid}: In this condition, 26 participants reported that their perceptions would not have changed had the correction been shown at a different time, while 13 participants indicated that their perceptions would. One participant was uncertain of how they would have reacted to an alternative timing. Participants who believed that timing would have made a difference frequently criticised the mid-placement for disrupting the flow of content and making the



(a)



(b)

Figure 6: Predicted probabilities of perceived credibility reduction across different predictors. Plot (a) shows the effect of debunking time, indicating significantly higher reductions when the correction appears at the end compared to the middle. Plot (b) shows a negative association between age and probability of perceived credibility, with younger participants showing higher probability of belief reduction when exposed to correction video.

correction feel intrusive. One participant noted, *"... I really disliked how the debunking video interrupted the video before it finished. It made me more annoyed at the intruding video."* ($P59_{mid}$). Another elaborated, *"I think the start or end would have made it feel more balanced, but interrupting in the middle devalued the presentation* *that was already going on since it suggested this needed to be corrected immediately and wasn't worth listening to all the way through."* ($P73_{mid}$). Participants who preferred alternative timings highlighted distinct advantages in both start and end placement. End placement was valued for allowing viewers to fully process the misinformation claim before comparing it with corrective content, as one explained: *"The end of the video is probably the best place. This way we have all the info from the original video and then can compare it to the information in the debunking video."* ($P55_{mid}$). By contrast, start placement was seen as helpful for priming a more critical mindset during exposure to misinformation, with one participant reflecting, *"If I saw the debunking video before the first one, I may think about it more during the first one and be more critical of it."* ($P49_{mid}$).

4.2.3 T_{start}: In this condition, 26 participants indicated that their perceptions would not have changed had the correction been shown at a different time. A few participants (n = 3) stated that they were unsure how their perceptions might have changed, while 1 participant did not clearly indicate a preference. A smaller number of participants (n=10), however, suggested that alternative placements might have been more effective. Start placement was considered effective for offering an immediate prompt to approach the content critically. One participant who favored this timing explained, *"no its good in the beginning to let you know to be skeptical even if you don't watch it"* ($P2_{start}$). Some disliked the start placement because it made them focus more on the misinformation video, as one reflected: *"... the first video actually irritated me, and made me more interested in hearing the second opinion."* ($P6_{start}$). End placement was viewed to offer a stronger impact, described as feeling more "mind-blowing" after watching the misinformation video in full; *"The debunking videos might have more impact if shown at the end because it felt more like a mind blowing video."* ($P9_{start}$).

5 Discussion

Short-form videos have emerged as an increasingly powerful vehicle for misinformation due to their ease of creation and distribution, as well as their highly persuasive nature. Traditional moderation approaches, such as warning labels or fact-checking links, often prove ineffective in these fast-paced environments where subtle cues are easily ignored. Video-based corrections present a promising alternative, as they deliver factual content within the same persuasive medium as the misinformation [37]. Yet, the impact of such corrections may depend on their timing, as corrections and misinformation present competing narratives, the order of exposure can shape how viewers interpret and remember the information. Therefore, we systematically evaluated three timing conditions to examine whether presenting corrections before, during, or after exposure to misinformation differentially affects users' perceived credibility of false claims.

5.1 Timing Does Matter When Correcting Video Misinformation

Confirming prior work on the effectiveness of correction videos [7, 37], our results also show that presenting a correction video reduces belief in misinformation. However, our findings show that the timing of the correction plays a critical role in determining its effectiveness. We found that corrections were most effective at the T_{end} condition and least effective at T_{mid}. In other words, participants who viewed the correction video *after* the misinformation video were more likely to reduce their initial belief in the misinformation. By contrast, when the same correction appeared *during* the misinformation video, participants were less likely to update their beliefs, even though some reduction still occurred.

Participants favoured the T_{end} condition because viewing the correction after the misinformation allowed them to fully process the false content before evaluating it against the corrective information. As prior research suggests, this positions the correction as a form of feedback, which can facilitate learning [44]. Moreover, prediction-error, the discrepancy that arises when new information contradicts existing expectations, has been shown to enhance the learning of novel facts [95]. In addition, the *hypercorrection effect*, which refers to the phenomenon that high-confidence errors are more likely to be corrected than low-confidence errors [17], may also have played a role in our findings. In our study, corrections were applied to misinformation that participants had initially rated as highly believable. As a result, when corrective feedback contradicted these strongly held beliefs, the element of surprise may have promoted deeper encoding of the corrective information and facilitated stronger belief revision [32].

In the least effective condition, T_{mid}, participants were interrupted during the misinformation video and shown corrective information, encouraging them to reflect on the accuracy of what they were watching. Prior studies suggest that slowing people down while they consume information can improve their ability to spot misinformation [90, 102]. This is because false beliefs are often linked to limited use of 'System 2 thinking', the slower, more deliberate mode of reasoning [91, 93]. Giving people a chance to pause and reflect is one way to trigger this deeper processing. However, our results show that T_{mid} was the least effective timing condition. Video narratives can strongly engage viewers by capturing their attention, emotions, and imagination. When a video is stopped midway, this immersive experience is disrupted. Unlike text, which lacks the same level of continuous narrative flow so a pause can provide space for reflection, interruptions in video break the flow of the story. Such disruptions may feel confusing or irritating to viewers, as their experience is cut short, which in turn can make them less receptive to corrective information, particularly when it challenges their initial beliefs [103].

Prior work on the timing of corrections for text-based misinformation has produced mixed results. Some studies report that corrections are most effective when delivered after exposure [12, 24], while others suggest that timing has little to no effect on corrective impact [97, 98]. While our findings align with earlier work, it is important to note that these studies on correction timing were conducted on political misinformation. Brashier et al. [12] found that the effectiveness of corrections did not differ between the *during* and *after* conditions based on participants' partisan alignment. They also reported that presenting the correction after the misinformation was more effective than presenting it before or during, even when the misinformation was politically aligned with the participant. However, although not focused on correction timing specifically, prior work has shown that corrections can be less effective when they challenge a strongly held belief or require people to change their attitudes [96, 115]. Political misinformation behaves differently from apolitical misinformation as it often connects strongly to people's existing beliefs [8] and partisan identities [35, 57]. This makes political misinformation more persistent and harder to correct [58, 81]. Given these dynamics, the political nature of the content may help explain the inconsistent results observed in text-based correction studies that examine timing.

As our study focuses on apolitical misinformation, we cannot assume that the same timing effects would apply to political content. Future work should directly compare political and apolitical misinformation in video-based settings to examine whether political identity changes how timing influences the impact of corrections, and whether different timing strategies may be needed for political versus non-political misinformation.

5.2 Recommendations for Misinformation Correction Efforts on Short-Form Video Platforms

Our findings reinforce that providing a correction for misinformation, regardless of when it is delivered, is always preferable to providing none at all. However, we also demonstrate that timing plays a critical role: corrections were most effective when shown *after* exposure to misinformation, and least effective when shown *during* it. This suggests that while debunking videos can be valuable interventions, their impact may be maximised when platforms are able to control when users encounter them. This raises a central design question: **how can platforms ensure that debunking content is delivered at the most effective moment to counter misinformation?** At present, debunking videos typically circulate independently, meaning that users might only encounter them incidentally, sometimes days later, or even not at all, long after the misinformation has already shaped their beliefs. This temporal gap significantly undermines their corrective potential.

One practical solution is to directly attach corrections to misinformation videos, presenting them *after* the misinformation video has finished playing. This could be achieved using a design similar to our study, where an overlay appeared after the video to notify users of potentially misleading content and offered the option to view a debunking video. On real-world platforms, debunking videos could be either algorithmically- or community-sourced. The design approach of using an overlay addresses limitations of current moderation strategies. Warning labels are often vague and shift effort to users, making them less effective against the emotional impact of misinformation [30]. External fact-check links also disrupt browsing and discourage engagement [38]. In contrast, overlays embed corrections in the same modality, providing immediate, persuasive alternatives while keeping users on-platform with minimal disruption.

> **Recommendation 1:** Platforms could implement post-exposure correction delivery by attaching corrective content directly to misinformation videos, ensuring users encounter debunking information immediately *after* the misleading content ends.

Another approach is in-feed pairing, where debunking videos are algorithmically surfaced immediately after or alongside the misinformation video in the content queue. This design could mimic the "next recommended video" structure but deliberately prioritise corrective content rather than leaving sequencing to chance. Algorithmic nudging has proven to increase both recommendations for, and consumption of, reliable news without reducing user engagement [128]. Furthermore, such algorithmic adjustments have been found to decrease belief in misinformation and promote a more balanced information diet [68, 77]. This approach allows users to encounter accurate information organically while, similar to the overlay design, keeping them within the platform and reducing the effort required to seek out corrective content elsewhere.

> **Recommendation 2:** Platforms could implement in-feed corrective sequencing by algorithmically positioning debunking videos directly after the misinformation videos within the content feed.

However, these platform-driven approaches raise challenges around autonomy and trust. Users may resist corrections that appear to be imposed top-down, especially given widespread skepticism toward fact-checking organisations. A way forward is to leverage the wisdom of the crowds [114]. Social media architectures can be adapted to incorporate user-generated signals of information quality, giving legitimacy to corrections through collective input rather than institutional authority. Platforms like X have implemented Community Notes[4], which allow users to collaboratively add context to misleading posts, while Meta and Reddit have experimented with crowdsourced credibility evaluations [20, 49]. These initiatives demonstrate that corrections sourced from peers are often more persuasive and better received than platform-level or fact-checker interventions.

> **Recommendation 3:** Platforms could enable crowdsourced corrections by allowing users to easily flag misleading content, upvote trusted debunking videos, and reward creators whose corrections receive strong community support. Platforms can then aggregate these signals to prioritise and surface high-quality corrective videos immediately after misinformation exposure.

A similar practice, however, has yet to be systematically implemented in short-form video platforms. Given the rise of debunking videos, users are already engaging in corrective practices by producing and sharing counter-content. Platforms do not need to create these interventions from scratch; instead, they can provide a "small push" by curating, aggregating, and strategically surfacing debunking videos at the most effective moment; particularly after misinformation exposure. These strategies can help platforms integrate timing into intervention design, ensuring that corrections are both effective and trusted.

5.3 Limitations & Future Work

There are several limitations to this study. First, we focused exclusively on short-form video corrections and their temporal effects. The dynamics may differ for long-form videos, where information volume and consumption patterns are substantially different. Future research should investigate the temporal effects of long-form content. Second, our analysis was limited to scientific misinformation. Other domains, such as political or social misinformation, may trigger distinct audience reactions. More research is needed to assess

[4] https://communitynotes.x.com/guide/en/about/introduction

the extent to which our findings generalise to these other domains. Third, participants viewed videos sequentially and with more attention than typical short-form video use, which generally occurs in a rapid and often distracted state. This design choice was necessary to enhance experimental control. Achieving high ecological realism is difficult without direct access to commercial platforms, access available to only a few researchers. Future research should further explore the timing dimension of short-form video consumption in more naturalistic interaction environments to strengthen ecological validity. Fourth, while we used lesser-known creators to avoid familiarity biases, real-world misinformation often spreads through celebrities and influencers [80]. Future work should examine how corrections function in these contexts, particularly given parasocial relationships [48], where audiences form trust and attachment to well-known figures. These dynamics could influence how and *when* corrections are received and should be explored to extend our findings beyond the context of unknown creators. Fifth, the misinformation topics used in this study were intentionally selected to be relatively benign to establish a controlled foundation for examining correction timing without the confounding influence of strong emotional responses. However, real misinformation ecosystems are dominated by politically and emotionally charged content [59], which can evoke high emotionality [11] and reduce analytical processing [73]. Future research should investigate whether correction timing operates differently when users are exposed to high-arousal misinformation. Sixth, since our objective was to examine the relative effectiveness of correction placement, we did not include a no-correction control condition. Future work could incorporate a no-correction baseline to assess absolute effect sizes and further clarify the temporal impact of corrections. Seventh, our study drew only from a U.S.-based participant pool, which limits the cultural and linguistic diversity of our findings. Exploring this topic in other countries and languages would provide a richer understanding of how users with different cultural backgrounds engage with such platforms. Finally, we measured pre-post credibility shift in line with prior research on misinformation [7, 37]. However, the durability of video-based corrections remains an open question. While work in learning suggests that delayed feedback can sometimes be more effective than immediate feedback [16], studies on text-based corrections show mixed results [18, 33, 98]. Future research should therefore examine whether corrective effects in video contexts persist over time and identify the conditions under which they may fade or endure.

6 Conclusion

Video-based corrections have consistently demonstrated effectiveness in countering misinformation. However, little attention has been given to their temporal effects, specifically when the correction should be presented: before, during, or after exposure to misinformation. This dimension is crucial, as the order of information, particularly when conflicting claims are involved, can shape how audiences process and retain content. In our study, we systematically examined correction timing and found that while presenting a correction is always better than not presenting one, corrections shown *after* exposure to misinformation were most effective in reducing misperceptions. In contrast, corrections presented *during* exposure were least effective. These findings offer valuable insights: platforms could directly attach correction videos to follow misinformation content, or leverage recommendation algorithms to ensure corrective content is surfaced immediately afterward. Embedding timing into intervention design would enable platforms to more effectively mitigate the spread of misinformation on short-form video platforms.

References

[1] Sara Abdali, Rutuja Gurav, Siddharth Menon, Daniel Fonseca, Negin Entezari, Neil Shah, and Evangelos E Papalexakis. 2021. Identifying misinformation from website screenshots. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. 2–13.

[2] Sara Abdali, Sina Shaham, and Bhaskar Krishnamachari. 2024. Multi-modal misinformation detection: Approaches, challenges and opportunities. *Comput. Surveys* 57, 3 (2024), 1–29.

[3] Nadia Alonso-López, Pavel Sidorenko-Bautista, Fábio Giacomelli, et al. 2021. Beyond challenges and viral dance moves: TikTok as a vehicle for disinformation and fact-checking in Spain, Portugal, Brazil, and the USA. *Anàlisi* (2021), 65–84.

[4] Scott Appling, Amy Bruckman, and Munmun De Choudhury. 2022. Reactions to Fact Checking. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–17.

[5] David S Axelrod, Brian P Harper, and John C Paolillo. 2022. YouTube COVID-19 vaccine misinformation on Twitter: platform interactions and moderation blind spots. *arXiv preprint arXiv:2208.13000* (2022).

[6] Michael Berkebile-Weinberg, Danielle Goldwert, Kimberly C Doell, Jay J Van Bavel, and Madalina Vlasceanu. 2024. The differential impact of climate interventions along the political divide in 60 countries. *Nature communications* 15, 1 (2024), 3885.

[7] Puneet Bhargava, Katie MacDonald, Christie Newton, Hause Lin, and Gordon Pennycook. 2023. How effective are TikTok misinformation debunking videos? *Harvard Kennedy School Misinformation Review* (2023).

[8] Leticia Bode and Emily K Vraga. 2015. In related news, that was wrong: The correction of misinformation through related stories functionality in social media. *Journal of Communication* 65, 4 (2015), 619–638.

[9] Leticia Bode and Emily K Vraga. 2018. See something, say something: Correction of global health misinformation on social media. *Health communication* 33, 9 (2018), 1131–1140.

[10] Leticia Bode, Emily K Vraga, and Melissa Tully. 2020. Do the right thing: Tone may not affect correction of misinformation on social media. *Harvard Kennedy School Misinformation Review* (2020).

[11] William J Brady, Julian A Wills, John T Jost, Joshua A Tucker, and Jay J Van Bavel. 2017. Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences* 114, 28 (2017), 7313–7318.

[12] Nadia M Brashier, Gordon Pennycook, Adam J Berinsky, and David G Rand. 2021. Timing matters when correcting fake news. *Proceedings of the National Academy of Sciences* 118, 5 (2021), e2020043118.

[13] Nadia M Brashier and Daniel L Schacter. 2020. Aging in an era of fake news. *Current directions in psychological science* 29, 3 (2020), 316–323.

[14] Yuyan Bu, Qiang Sheng, Juan Cao, Peng Qi, Danding Wang, and Jintao Li. 2023. Combating online misinformation videos: Characterization, detection, and future directions. In *Proceedings of the 31st ACM International Conference on Multimedia*. 8770–8780.

[15] Cody Buntain, Richard Bonneau, Jonathan Nagler, and Joshua A Tucker. 2021. YouTube recommendations and effects on sharing across online social platforms. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–26.

[16] Andrew C Butler, Jeffrey D Karpicke, and Henry L Roediger III. 2007. The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied* 13, 4 (2007), 273.

[17] Brady Butterfield and Janet Metcalfe. 2001. Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27, 6 (2001), 1491.

[18] Je Hoon Chae, Tim Groeling, and Hyunjin Song. 2025. Time is the fire in which message effects burn: Decay and sustenance of correction effects over time. *Journal of Communication* (2025), jqaf030.

[19] James M Clark and Allan Paivio. 1991. Dual coding theory and education. *Educational psychology review* 3, 3 (1991), 149–210.

[20] Giulio Corsi, Elizabeth Seger, et al. 2024. Crowdsourcing the Mitigation of disinformation and misinformation: The case of spontaneous community-based moderation on Reddit. *Online Social Networks and Media* 43 (2024), 100291.

[21] Michael Craig and Santosh Vijaykumar. 2023. One dose is not enough: the beneficial effect of corrective COVID-19 information is diminished if followed by misinformation. *Social Media+ Society* 9, 2 (2023), 20563051231161298.

[22] Joshua Cuevas and Bryan L Dawson. 2018. A test of two alternative cognitive processing models: Learning styles and dual coding. *Theory and Research in Education* 16, 1 (2018), 40–64.

[23] Richard L Daft and Robert H Lengel. 1986. Organizational information requirements, media richness and structural design. *Management science* 32, 5 (1986), 554–571.

[24] Yue Dai, Wenting Yu, and Fei Shen. 2021. The effects of message order and debiasing information in misinformation correction. *International Journal of Communication* 15 (2021), 21.

[25] Meredith E David and James A Roberts. 2024. TikTok brain: An investigation of short-form video use, self-control, and phubbing. *Social Science Computer Review* (2024), 08944393241279422.

[26] Emily Denniss and Rebecca Lindberg. 2025. Social media and the spread of misinformation: infectious and a threat to public health. *Health promotion international* 40, 2 (2025), daaf023.

[27] Regional Development Department of Infastructure, Transport and Australia Communications. 2020. Fact or fiction? The truth about 5G. www.infrastructure.gov.au/sites/default/files/fact-or-fiction-the-truth-about-5g.pdf.

[28] Olga Dontcheva-Navratilova. 2023. Persuasion in multimodal digital genres: Building credibility in video abstracts. *ESP Today* 11, 2 (2023), 213–236.

[29] Ullrich KH Ecker, Stephan Lewandowsky, and Matthew Chadwick. 2020. Can corrections spread misinformation to new audiences? Testing for the elusive familiarity backfire effect. *Cognitive Research: Principles and Implications* 5, 1 (2020), 41.

[30] Ullrich KH Ecker, Stephan Lewandowsky, and David TW Tang. 2010. Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & cognition* 38 (2010), 1087–1100.

[31] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (2007), 175–191.

[32] Lisa K Fazio and Elizabeth J Marsh. 2009. Surprising feedback improves later memory. *Psychonomic Bulletin & Review* 16, 1 (2009), 88–92.

[33] R Kelly Garrett and Brian E Weeks. 2013. The promise and peril of real-time corrections to political misperceptions. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 1047–1058.

[34] Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

[35] Tim Groeling. 2013. Media bias by the numbers: Challenges and opportunities in the empirical study of partisan news. *Annual Review of Political Science* 16, 1 (2013), 129–151.

[36] Gisel Bastidas Guacho, Sara Abdali, Neil Shah, and Evangelos E Papalexakis. 2018. Semi-supervised content-based detection of misinformation via tensor embeddings. In *2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*. IEEE, 322–325.

[37] Suwani Gunasekara, Saumya Pareek, Ryan M Kelly, and Jorge Goncalves. 2025. The Influence of Content Modality on Perceptions of Online Misinformation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–10.

[38] Chen Guo, Nan Zheng, and Chengqi Guo. 2023. Seeing is not believing: a nuanced view of misinformation warning efficacy on video-sharing social media platforms. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–35.

[39] Saqib Hakak, Mamoun Alazab, Suleman Khan, Thippa Reddy Gadekallu, Praveen Kumar Reddy Maddikunta, and Wazir Zada Khan. 2021. An ensemble machine learning approach through effective feature extraction to classify fake news. *Future Generation Computer Systems* 117 (2021), 47–58.

[40] Aniko Hannak, Drew Margolin, Brian Keegan, and Ingmar Weber. 2014. Get back! you don't know me like that: The social mediation of fact checking interventions in twitter conversations. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 8. 187–196.

[41] Katrin Hartwig, Tom Biselli, Franziska Schneider, and Christian Reuter. 2024. 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*. 1–20.

[42] Katrin Hartwig, Frederic Doell, and Christian Reuter. 2024. The landscape of user-centered misinformation interventions-a systematic literature review. *Comput. Surveys* 56, 11 (2024), 1–36.

[43] Amelia Hassoun, Ian Beacock, Sunny Consolvo, Beth Goldberg, Patrick Gage Kelley, and Daniel M Russell. 2023. Practicing information sensibility: How Gen Z engages with online information. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[44] John Hattie and Helen Timperley. 2007. The power of feedback. *Review of educational research* 77, 1 (2007), 81–112.

[45] Zhiyu He, Shaorun Zhang, Peijie Sun, Jiayu Li, Xiaohui Xie, Min Zhang, and Yiqun Liu. 2023. Understanding user immersion in online short video interaction. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 731–740.

[46] Alex Hern. 2021. TikTok to introduce warnings on content to help tackle misinformation — theguardian.com. https://www.theguardian.com/technology/2021/feb/04/tiktok-to-introduce-warnings-on-content-to-help-tackle-misinformation.

[47] Benjamin Horne and Sibel Adali. 2017. This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *Proceedings of the international AAAI conference on web and social media*, Vol. 11. 759–766.

[48] Donald Horton and R Richard Wohl. 1956. Mass communication and para-social interaction: Observations on intimacy at a distance. *psychiatry* 19, 3 (1956), 215–229.

[49] https://www.abc.net.au/news/kevin nguyen/10373694. 2025. Meta opts to leave fact-checking to unpaid users. What could go wrong? https://www.abc.net.au/news/2025-01-08/elon-musk-community-notes-on-meta-facebook-explained/104794984.

[50] Emelia May Hughes, Renee Wang, Prerna Juneja, Tony W Li, Tanushree Mitra, and Amy X Zhang. 2024. Viblio: Introducing credibility signals and citations to video-sharing platforms. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–20.

[51] Minyoung Huh, Andrew Liu, Andrew Owens, and Alexei A Efros. 2018. Fighting fake news: Image splice detection via learned self-consistency. In *Proceedings of the European conference on computer vision (ECCV)*. 101–117.

[52] Eslam Hussein, Prerna Juneja, and Tanushree Mitra. 2020. Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (May 2020), 1–27. doi:10.1145/3392854

[53] Sylvia Iasulaitis, Alan Demetrius Baria Valejo, and Leonardo Ribeiro dos Santos. 2025. The influence of partisanship on the roots of anti-vaccine attitudes in the context of the COVID-19 pandemic. *Vaccine: X* (2025), 100707.

[54] Md Rafiqul Islam, Shaowu Liu, Xianzhi Wang, and Guandong Xu. 2020. Deep learning for misinformation detection on online social networks: a survey and new perspectives. *Social Network Analysis and Mining* 10, 1 (2020), 82.

[55] isolomons. 2021. How We're Tackling Misinformation Across Our Apps — about.fb.com. https://about.fb.com/news/2021/03/how-were-tackling-misinformation-across-our-apps/.

[56] Caroline Jack. 2017. Lexicon of lies: Terms for problematic information. *Data & Society* 3, 22 (2017), 1094–1096.

[57] Jennifer Jerit and Jason Barabas. 2012. Partisan perceptual bias and the information environment. *The Journal of Politics* 74, 3 (2012), 672–684.

[58] Jennifer Jerit and Yangzi Zhao. 2020. Political misinformation. *Annual Review of Political Science* 23, 1 (2020), 77–94.

[59] Shan Jiang and Christo Wilson. 2018. Linguistic signals under misinformation and fact-checking: Evidence from user comments on social media. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–23.

[60] Paige L Kemp, Vanessa M Loaiza, Colleen M Kelley, and Christopher N Wahlheim. 2024. Correcting fake news headlines after repeated exposure: memory and belief accuracy in younger and older adults. *Cognitive Research: Principles and Implications* 9, 1 (2024), 55.

[61] Panayiota Kendeou, Reese Butterfuss, Jasmine Kim, and Martin Van Boekel. 2019. Knowledge revision through the lenses of the three-pronged approach. *Memory & Cognition* 47, 1 (2019), 33–46.

[62] Namit Khurana, Mohammad Majid Akhtar, Rahat Masood, Salil S Kanhere, and Benjamin Turnbull. 2025. Navigating the Gray Interventions: The Impact of Soft Moderation Techniques Across Social Media Platforms. In *Companion Proceedings of the ACM on Web Conference 2025*. 1067–1071.

[63] Young Kim and Hyunji Lim. 2023. Debunking misinformation in times of crisis: Exploring misinformation correction strategies for effective internal crisis communication. *Journal of Contingencies and Crisis Management* 31, 3 (2023), 406–420.

[64] Young Kim and Hyunji Dana Lim. 2025. "Truth Sandwich" To Dispel Crisis Misinformation: Exploring the Separate and Combined Impacts of Inoculation, Narratives, and Stealing Thunder on Correcting Crisis Misinformation on Social Media. *Mass Communication and Society* (2025), 1–21.

[65] Nils C Köbis, Barbora Doležalová, and Ivan Soraperra. 2021. Fooled twice: People cannot detect deepfakes but think they can. *Iscience* 24, 11 (2021).

[66] Anthony L. Komaroff. 2019. Ask the doctor: Microwave's impact on food - Harvard Health — health.harvard.edu. https://www.health.harvard.edu/staying-healthy/ask-the-doctor-microwaves-impact-on-food.

[67] Johannes Kotz, Helge Giese, and Laura M König. 2023. How to debunk misinformation? An experimental online study investigating text structures and headline formats. *British Journal of Health Psychology* 28, 4 (2023), 1097–1112.

[68] Jana Lasser and Nikolaus Poechhacker. 2025. Designing social media content recommendation algorithms for societal good. *Annals of the New York Academy of Sciences* (2025).

[69] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096.

[70] Stephan Lewandowsky, John Cook, Ullrich Ecker, Dolores Albarracin, Michelle A Amazeen, Panayiota Kendou, Doug Lombardi, Eryn Newman, Gordon Pennycook, Ethan Porter, et al. 2020. *The debunking handbook 2020.*

[71] Stephan Lewandowsky, Ullrich KH Ecker, and John Cook. 2017. Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of applied research in memory and cognition* 6, 4 (2017), 353–369.

[72] Ryan Hardesty Lewis and Alex Conway. 2024. Streaming the Future of Sustainability. *ACM SIGENERGY Energy Informatics Review* 4, 5 (2024), 166–171.

[73] Ming-Hui Li, Zhiqin Chen, and Li-Lin Rao. 2022. Emotion, analytic thinking and susceptibility to misinformation during the COVID-19 outbreak. *Computers in Human Behavior* 133 (2022), 107295.

[74] Li-Chao Ma, Juan Liu, Cui-Liu Jiao, Shao-Ying Du, Ruo-Lan Zhang, Xiao-Juan Ding, Xiao-Yang Shi, and Yan Wang. 2024. Exploring the effects of an online learning platform in stage III cardiac rehabilitation for individuals with coronary heart disease: Randomized controlled study. *Medicine* 103, 37 (2024), e39497.

[75] Drew B Margolin, Aniko Hannak, and Ingmar Weber. 2018. Political fact-checking on Twitter: When do corrections have an effect? *Political communication* 35, 2 (2018), 196–219.

[76] Cameron Martel, Mohsen Mosleh, and David G Rand. 2021. You're definitely wrong, maybe: Correction style has minimal effect on corrections of misinformation online. *Media and Communication* 9, 1 (2021), 120–133.

[77] J Nathan Matias. 2023. Influencing recommendation algorithms to reduce the spread of unreliable news by encouraging humans to fact-check articles, in a field experiment. *Scientific Reports* 13, 1 (2023), 11715.

[78] Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica* 22, 3 (2012), 276–282.

[79] Dar Meshi and Maria D Molina. 2025. Problematic social media use is associated with believing in and engaging with fake news. *PloS one* 20, 5 (2025), e0321361.

[80] Dan Milmo. 2024. Online influencers need 'urgent' fact-checking training, warns Unesco — theguardian.com. https://www.theguardian.com/media/2024/nov/26/online-influencers-need-urgent-fact-checking-training-warns-unesco.

[81] David S Morris, Jonathan S Morris, and Peter L Francia. 2020. A fake news inoculation? Fact checkers, partisan identification, and the power of misinformation. *Politics, Groups, and Identities* 8, 5 (2020), 986–1005.

[82] ABC Radio National. 2024. MSG's messy myths and racist tropes. www.abc.net.au/listen/programs/whats-that-rash/msg-additive-myths/104447016.

[83] Shuo Niu, Zhicong Lu, Amy X Zhang, Jie Cai, Carla F Griggio, and Hendrik Heuer. 2023. Building Credibility, Trust, and Safety on Video-Sharing Platforms. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–7.

[84] Heather L O'Brien, Nilou Davoudi, and Melissa Nelson. 2025. TikTok as information space: A scoping review of information behavior on TikTok. *Library & Information Science Research* 47, 4 (2025), 101379.

[85] The Editors of Encyclopædia Britannica. [n. d.]. Are There Really Right-Brained and Left-Brained People? | Britannica — britannica.com. https://www.britannica.com/story/are-there-really-right-brained-and-left-brained-people.

[86] University of South Australia. 2025. Myth busted: Healthy habits take longer than 21 days to set in — unisa.edu.au. unisa.edu.au/media-centre/Releases/2025/myth-busted-healthy-habits-take-longer-than-21-days-to-set-in/. [Accessed 06-09-2025].

[87] Daniel M Oppenheimer, Tom Meyvis, and Nicolas Davidenko. 2009. Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of experimental social psychology* 45, 4 (2009), 867–872.

[88] Kostantinos Papadamou, Savvas Zannettou, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, and Michael Sirivianos. 2022. "It is just a flu": assessing the effect of watch history on YouTube's pseudoscientific video recommendations. In *Proceedings of the international AAAI conference on web and social media*, Vol. 16. 723–734.

[89] Gordon Pennycook, James Allan Cheyne, Nathaniel Barr, Derek J Koehler, and Jonathan A Fugelsang. 2015. On the reception and detection of pseudo-profound bullshit. *Judgment and Decision making* 10, 6 (2015), 549–563.

[90] Gordon Pennycook, Ziv Epstein, Mohsen Mosleh, Antonio A Arechar, Dean Eckles, and David G Rand. 2021. Shifting attention to accuracy can reduce misinformation online. *Nature* 592, 7855 (2021), 590–595.

[91] Gordon Pennycook and David G Rand. 2019. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* 188 (2019), 39–50.

[92] Gordon Pennycook and David G Rand. 2020. Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of personality* 88, 2 (2020), 185–200.

[93] Gordon Pennycook and David G Rand. 2021. The psychology of fake news. *Trends in cognitive sciences* 25, 5 (2021), 388–402.

[94] Stefan Pfattheicher and Simon Schindler. 2016. Misperceiving bullshit as profound is associated with favorable views of Cruz, Rubio, Trump and conservatism. *PloS one* 11, 4 (2016), e0153419.

[95] Alex Pine, Noa Sadeh, Aya Ben-Yakov, Yadin Dudai, and Avi Mendelsohn. 2018. Knowledge acquisition is governed by striatal prediction errors. *Nature communications* 9, 1 (2018), 1673.

[96] Clara Pretus, Camila Servin-Barthet, Elizabeth A Harris, William J Brady, Oscar Vilarroya, and Jay J Van Bavel. 2023. The role of political devotion in sharing partisan misinformation and resistance to fact-checking. *Journal of Experimental Psychology: General* 152, 11 (2023), 3116.

[97] Toby Prike and Ullrich KH Ecker. 2023. Effective correction of misinformation. *Current Opinion in Psychology* 54 (2023), 101712.

[98] Patrick R Rich and Maria S Zaragoza. 2020. Correcting misinformation in news stories: An investigation of correction timing and correction durability. *Journal of Applied Research in Memory and Cognition* 9, 3 (2020), 310–322.

[99] Daniel C Richardson, Nicole K Griffin, Lara Zaki, Auburn Stephenson, Jiachen Yan, Thomas Curry, Richard Noble, John Hogan, Jeremy I Skipper, and Joseph T Devlin. 2020. Engagement in video and audio narratives: Contrasting self-report and physiological measures. *Scientific Reports* 10, 1 (2020), 11298.

[100] Jon Roozenbeek, Sander van der Linden, and Thomas Nygren. 2020. Prebunking interventions based on the psychological theory of "inoculation" can reduce susceptibility to misinformation across cultures. *Harvard Kennedy School Misinformation Review* (Feb. 2020). doi:10.37016//mr-2020-008

[101] Thomas J Rudolph and Marc J Hetherington. 2021. Affective polarization in political and nonpolitical settings. *International Journal of Public Opinion Research* 33, 3 (2021), 591–606.

[102] Nikita A Salovich and David N Rapp. 2021. Misinformed and unaware? Metacognition and the influence of inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 47, 4 (2021), 608.

[103] Angeline Sangalang, Yotam Ophir, and Joseph N Cappella. 2019. The potential for narrative correctives to combat misinformation. *Journal of communication* 69, 3 (2019), 298–319.

[104] Filipo Sharevski, Jennifer Vander Loop, Peter Jachim, Amy Devine, and Sanchari Das. 2024. 'Debunk-It-Yourself': Health Professionals Strategies for Responding to Misinformation on TikTok. In *Proceedings of the New Security Paradigms Workshop*. 35–55.

[105] Filipo Sharevski and Aziz N Zeidieh. 2023. "I Just Didn't Notice It:" Experiences with Misinformation Warnings on Social Media amongst Users Who Are Low Vision or Blind. In *Proceedings of the 2023 New Security Paradigms Workshop*. 17–33.

[106] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter* 19, 1 (2017), 22–36.

[107] Kai Shu, Suhang Wang, and Huan Liu. 2019. Beyond news contents: The role of social context for fake news detection. In *Proceedings of the twelfth ACM international conference on web search and data mining*. 312–320.

[108] Keith E Stanovich and Maggie E Toplak. 2023. Actively open-minded thinking and its measurement. *Journal of Intelligence* 11, 2 (2023), 27.

[109] Daniel Stockemer and Jean Nicolas Bordeleau. 2024. Understanding climate change conspiracy beliefs: A comparative outlook. *Harvard Kennedy School Misinformation Review* 5, 6 (2024).

[110] Florian Stoeckel, Sabrina Stöckli, Besir Ceka, Chiara Ricchi, Ben Lyons, and Jason Reifler. 2024. Social corrections act as a double-edged sword by reducing the perceived accuracy of false and real news in the UK, Germany, and Italy. *Communications Psychology* 2, 1 (2024), 10.

[111] Evan Ogg Straub. 2025. Roundup on Research: The Myth of 'Learning Styles' | Online Teaching. https://onlineteaching.umich.edu/articles/the-myth-of-learning-styles/.

[112] Yan Su, Porismita Borah, and Xizhu Xiao. 2022. Understanding the "infodemic": Social media news use, homogeneous online discussion, self-perceived media literacy and misperceptions about COVID-19. *Online Information Review* 46, 7 (2022), 1353–1372.

[113] S Shyam Sundar, Maria D Molina, and Eugene Cho. 2021. Seeing is believing: Is video modality more powerful in spreading fake news via online messaging apps? *Journal of Computer-Mediated Communication* 26, 6 (2021), 301–319.

[114] James Surowiecki. 2005. The wisdom of crowds/james surowiecki. *NY.: Anchor* 5 (2005).

[115] Mark W Susmann and Duane T Wegener. 2023. How attitudes impact the continued influence effect of misinformation: The mediating role of discomfort. *Personality and Social Psychology Bulletin* 49, 5 (2023), 744–757.

[116] Yuko Tanaka, Hiromi Arai, Miwa Inuzuka, Yoichi Takahashi, Minao Kukita, Ryuta Iseki, and Kentaro Inui. 2025. 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*. 1–18.

[117] Yuko Tanaka, Miwa Inuzuka, Hiromi Arai, Yoichi Takahashi, Minao Kukita, and Kentaro Inui. 2023. Who does not benefit from fact-checking websites? A psychological characteristic predicts the selective avoidance of clicking uncongenial facts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[118] Teleprompter.com team. 2025. Teleprompter.com | TikTok Marketing Statistics 2025 — teleprompter.com. https://www.teleprompter.com/blog/tiktok-marketing-statistics.

[119] David R Thomas. 2006. A general inductive approach for analyzing qualitative evaluation data. *American journal of evaluation* 27, 2 (2006), 237–246.
[120] TikTok. 2025. Countering harmful misinformation guide — tiktok.com. https://www.tiktok.com/safety/en/harmful-misinformation-guide.
[121] Dayane Fumiyo Tokojima Machado, Alexandre Fioravante de Siqueira, Natiely Rallo Shimizu, and Leda Gitahy. 2022. It-which-must-not-be-named: COVID-19 misinformation, tactics to profit from it and to evade content moderation on YouTube. *Frontiers in Communication* 7 (2022), 1037432.
[122] Matus Tomlein, Branislav Pecher, Jakub Simko, Ivan Srba, Robert Moro, Elena Stefancova, Michal Kompan, Andrea Hrckova, Juraj Podrouzek, and Maria Bielikova. 2021. An Audit of Misinformation Filter Bubbles on YouTube: Bubble Bursting and Recent Behavior Changes. In *Fifteenth ACM Conference on Recommender Systems*. ACM. doi:10.1145/3460231.3474241
[123] Emily K Vraga and Leticia Bode. 2020. Defining misinformation and understanding its bounded nature: Using expertise and evidence for describing misinformation. *Political Communication* 37, 1 (2020), 136–144.
[124] Chloe Wittenberg, Ben M Tappin, Adam J Berinsky, and David G Rand. 2021. The (minimal) persuasive advantage of political video over text. *Proceedings of the National Academy of Sciences* 118, 47 (2021), e2114388118.
[125] Dannagal G Young, Kathleen Hall Jamieson, Shannon Poulsen, and Abigail Goldring. 2018. Fact-checking effectiveness as a function of format and tone: Evaluating FactCheck. org and FlackCheck. org. *Journalism & Mass Communication Quarterly* 95, 1 (2018), 49–75.
[126] YouTube. [n. d.]. YouTube misinformation policies - How YouTube Works — youtube.com. https://www.youtube.com/intl/ALL_in//howyoutubeworks/our-commitments/fighting-misinformation/.
[127] YouTube. 2025. Navigating YouTube news and information - How YouTube Works — youtube.com. https://www.youtube.com/intl/ALL_au//howyoutubeworks/product-features/news-information/#fact-checks.
[128] Xudong Yu, Muhammad Haroon, Ericka Menchen-Trevino, and Magdalena Wojcieszak. 2024. Nudging recommendation algorithms increases news consumption and diversity on YouTube. *PNAS nexus* 3, 12 (2024), pgae518.

A Summary of the videos

Table 2: Summary of topics, veracity labels, durations (seconds), likes, views and number of participants who saw each correction video.

Topic	Veracity	Duration (seconds)	Likes	Views	Participant views
5G is harmful to human health	Myth	58	5	238	120
	Truth	60	32	988	16
MSG is an unsafe flavor enhancer that poses risks to overall health	Myth	63	2781	122.9K	120
	Truth	59	166	1.6K	36
Microwaves cause cancer and other serious health issues due to radiation	Myth	60	5321	210.9K	120
	Truth	69	677	NA	14
Individual learn better when they receive information in their preferred learning style	Myth	57	5	559	120
	Truth	61	156	172	32
People can be divided as left and right brain people	Myth	26	70	2.4K	120
	Truth	48	840	843	97
It take 21 days to form a habit	Myth	31	1260	36.1K	120
	Truth	55	25	582	45