Re-exposure to AI-Generated Images During Misinformation Corrections Reduces False Beliefs by Improving Memory for Corrections

Sean Guo[a], Sen Mu[a], Christopher N. Wahlheim[b] & Xiaoqing Hu[a, c]

[a] Department of Psychology, The University of Hong Kong, Hong Kong SAR, China

[b] Department of Psychology, University of North Carolina at Greensboro, North Carolina, United States of America

[c] HKU-Shenzhen Institute of Research and Innovation, Shenzhen, China

Author Note

Correspondence concerning this article should be addressed to Xiaoqing Hu (Email: xiaoqinghu@hku.hk) and Christopher N. Wahlheim (Email: cnwahlhe@uncg.edu)

Abstract

AI-generated images are increasingly used to spread false information online, but it remains unclear how repeating these images during corrections affects false beliefs. Re-exposing images could improve memory for corrections, but it could also make false information appear more realistic. Across three preregistered experiments (total N = 884), we examined how re-exposing people to AI-generated images during corrections affected their belief in false headlines and memory for corrections. Image re-exposure during corrections reduced false belief and improved memory for corrections compared with corrections without image re-exposure. These effects persisted after a one-week delay and even when the AI origin of the images was not disclosed. Better memory for corrections was associated with lower false belief. These findings show that re-exposing people to AI-generated images during corrections improves memory for corrections and suggest that this method may lead to more durable encoding of corrections within fact-checks encountered in everyday online contexts.

Keywords: misinformation, AI-generated images, picture superiority effect, memory, corrections

Re-exposure to AI-Generated Images During Misinformation Corrections Reduces False Beliefs by Improving Memory for Corrections

With the rapid development of multimodal generative artificial intelligence (AI), tools creating AI-generated visual misinformation (AIVM) providing vivid evidence for false claims have become widespread. Mass exposure to such content could negatively affect democratic discourse, public health, and voter attitudes (Diakopoulos & Johnson, 2021; Heley et al., 2022). Research is therefore needed to identify effective ways to correct misinformation accompanied by AIVM. It is currently unclear whether corrections are more effective when they re-expose people to the original false images. Fact-checking websites vary in this respect: Sites such as Snopes (Snopes, 2026) display the original false image alongside the correction, whereas others, such as Politifact (Politifact, 2026), hide the image unless users click into the article. To address this issue, we investigated how re-exposing people to AIVM during corrections affects belief in false headlines and memory for corrections, and whether potential re-exposure effects depend on the probative value of images (i.e., whether images provide evidence for the claims).

Many studies show the continued influence effect (CIE), in which people continue to rely on misinformation for subsequent reasoning even after reading corrections, with effects re-emerging or increasing over time (Ecker et al., 2022). One explanation is that memory for corrections declines faster than memory for the misinformation. When corrections are forgotten, misinformation may guide judgments maladaptively (Maertens et al., 2025; Swire-Thompson et al., 2023). Consistent with this account, correction formats that enhance encoding and later memory for corrections are more likely to produce durable reductions in false beliefs (Wahlheim et al., 2026). Thus, identifying correction formats that improve memory for corrections is critical for reducing the continued influence of misinformation.

One possibility is that re-exposing images during corrections improves memory for the correction itself. Evidence for this possibility comes from research showing that memory is often better for word-image than word-word pairs (Hockley & Bancroft, 2011), a phenomenon known as the picture superiority effect (Paivio, 1971). Because correcting misinformation requires associating a claim with corrective information, re-exposure to AIVM during corrections may strengthen these associations and improve subsequent memory for the correction (Ecker et al., 2011; Gilbert et al., 1990). Although the picture superiority effect has not been directly examined in corrections of misinformation, related work shows that visualizations such as graphs can increase correction efficacy (Wijnker et al., 2022), suggesting that visual information may improve correction efficacy by enhancing memory for corrections. Moreover, improved memory for corrections produced by reminder-based correction methods is associated with greater reductions in false belief (Kemp et al., 2022, 2024; Wellons & Wahlheim, 2025). Together, these findings suggest that re-exposure to AIVM during corrections may reduce false belief by improving memory for corrections.

However, a recent study showed that adding nonprobative images to misinformation corrections did not improve correction efficacy (Whitehead et al., 2025). Therefore, an alternative possibility is that re-exposure to images could have no effect on or even undermine correction efficacy. Because images often appear to reflect reality, people tend to trust visual evidence (Messaris & Abraham, 2001), and seeing a visual representation of an event may increase trust by activating the realism heuristic (Sundar, 2008; Sundar et al., 2021). Consistent with this idea, a meta-analysis found that statements paired with images are often judged as more truthful (Seo, 2020). More realistic and probative AIVM images, meaning images that provide direct visual evidence for a claim, are also associated with greater belief in misinformation (Guo, Zhong, et al., 2025). If re-exposure increases the perceived realism of misinformation, people may fail to accept corrections and encode them less effectively, which could reduce memory for the corrections (O'Rear & Radvansky, 2020). This account therefore predicts that re-exposure to AIVM, relative to when images are not re-presented, will impair memory for corrections and ultimately decrease correction efficacy over time.

The Present Study

To test these competing predictions, we conducted three experiments examining how re-exposure to AIVM during corrections affects belief in false headlines and memory for corrections. First, participants viewed news headlines paired with images, some of which were AIVM accompanying false claims. Second, they viewed fact-checks indicating the veracity of each headline. Some fact-checks re-presented the original images whereas others did not, allowing us to test image re-exposure effects. Third, we measured belief in the headlines and memory for the corrections to assess belief updating and its reliance on memory. In Experiments 1 and 2, these measures were deployed immediately after correction and after one week to examine the durability of corrections. In Experiments 2 and 3, we additionally manipulated whether images provided strong visual evidence for the claims to test whether re-exposure effects depend on image probity. Finally, Experiment 3 refined the design by removing mention of AI-generated imagery from the corrections, allowing us to verify that the earlier effects reflected image re-exposure rather than references to AI-generated imagery. If AIVM re-exposure improves associative memory for corrections, false belief should be lower when images are re-presented than when they are omitted. In contrast, if AIVM re-exposure increases the realism of misinformation, false belief should be higher when images are re-presented than omitted.

Research Transparency Statement

General Disclosures

Conflicts of interest: All authors declare no conflicts of interest. Funding: The research was supported by the seed grants of the Population, Human Behavior and Wellness Strategic Research Theme of the Faculty of Social Sciences, The University of Hong Kong, National Natural Science Foundation of China (No. 32171056), and General Research Fund (No. 17614922) of Hong Kong Research Grants Council. The funding sources had no role other than financial support. Artificial intelligence: Generative Artificial Intelligence (Midjourney, 2025) was used to synthesize stimuli for Experiments 2 and 3. Ethics: This research was approved by the Human Research Ethics Committee of the University of Hong Kong (EA210341).

Experiment One Disclosures

Preregistration: The hypotheses, methods and analyses were preregistered (https://osf.io/qbrs7). There were minor deviations from the preregistration (see Supplementary Materials Section A). Materials: Given that AI-generated images may be used to spread false information online, materials are only available upon request. Data: All primary data are publicly available (https://osf.io/t65xg/). Analysis scripts: Analysis scripts are publicly available (https://osf.io/t65xg/).

Experiment Two Disclosures

Preregistration: The hypotheses, methods and analyses were preregistered (https://osf.io/t2s65). There were minor deviations from the preregistration (see Supplementary Materials Section A). Materials: Given that AI-generated images may be used to spread false information online, materials are only available upon request. Data: All primary data are publicly available (https://osf.io/t65xg/). Analysis scripts: Analysis scripts are publicly available (https://osf.io/t65xg/).

Experiment Three Disclosures

Preregistration: The hypotheses, methods and analyses were preregistered (https://osf.io/ng9h2/). There were no deviations from the preregistration. Materials: Given that AI-generated images may be used to spread false information online, materials are only available upon request. Data: All primary data are publicly available (https://osf.io/t65xg/). Analysis scripts: Analysis scripts are publicly available (https://osf.io/t65xg/).

Experiment 1

Experiment 1 provides the first characterization of the effects of AIVM re-exposure during corrections on false beliefs, memory for corrections, and their association over time.

Method

For full materials, power analysis, and procedure details, refer to Supplementary Material Section B. To protect against automated responses, we enabled Qualtrics' bot detection functionality (reCAPTCHA), prevented multiple submissions, and included attention check questions.

Participants

After exclusions, the final sample included 148 participants (51 men, 92 women, and 5 unknown gender) from Prolific (M_{age} = 38.6, *SD* = 13.0, range = 18 to 80 years old). An a priori power analysis determined a target sample size of 200, and data collection continued until this target was reached. Participants currently resided in the U.S., spoke English as their first language, and had an approval rating of at least 90% on Prolific. Fifty-six participants were excluded because they provided the same belief rating across trials, did not complete all phases, failed attention or honesty checks, or attempted the survey multiple times. The sample characteristics were based on prior research on images in misinformation (Newman et al., 2015; Whitehead et al., 2025), and we used an online platform to access participants relevant to the study's focus on digital misinformation exposure.

Design

The experiment used a 2 (AIVM: re-exposure vs. no re-exposure) × 2 (Headline veracity: true vs. false) within-participants design. Participants provided belief ratings in three phases (initial, immediate post-correction, and delayed post-correction) and ratings of memory for corrections in two phases (immediate and delayed post-correction).

Materials

We used headlines and AI-generated images from our earlier studies (Guo, Swire-Thompson, et al., 2025; Guo, Zhong, et al., 2025), which contained news about animals, accidents, natural disasters, and strange phenomena. We avoided political and health-related topics in our headlines to minimize effects of prior attitudes on belief. In total, there were 20 false and 20 true headline-image pairs.

Procedure

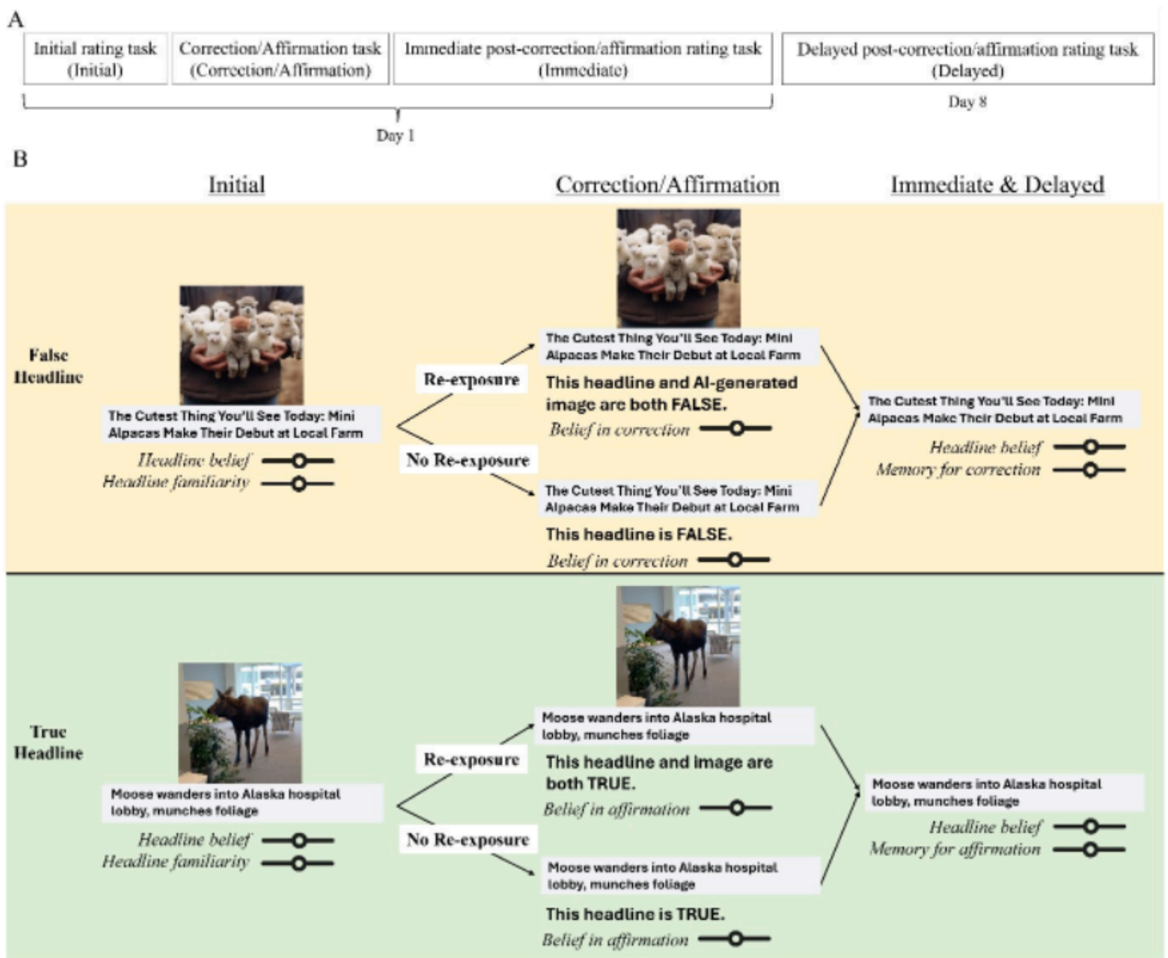
The experiment was administered using Qualtrics software (Qualtrics, Provo, UT). Figure 1 displays a procedural schematic and examples of experimental conditions. All ratings were self-paced, and participants clicked to advance to the next trial after each rating.

On the initial rating task, participants viewed 20 true headlines paired with real images and 20 false headlines paired with AIVM images. Each pair was randomly presented for 8 s. Upon viewing each pair, participants rated its veracity (1, definitely false, to 10, definitely true) and their familiarity with it (1, extremely unfamiliar, to 10, extremely familiar). Next, on the correction/affirmation task, participants viewed fact-checks that were corrections of false headlines and affirmations of true headlines from the prior phase. Participants were told to read and rate their belief in each correction or affirmation. Headlines were evenly and randomly split between the re-exposure and no re-exposure conditions, and were presented individually and randomly. In the re-exposure condition, the corresponding images were re-presented together with veracity labels indicating correction of false headlines ("this headline and AI-generated image are both FALSE") and affirmation of true headlines ("This headline and image are both TRUE"). In the no re-exposure condition, participants saw only veracity labels for the false headlines ("This headline is FALSE") and true headlines ("This headline is TRUE"). All fact-checks appeared for 8 s each. Participants then rated their belief that the fact-check was true from 1 (definitely not) to 10 (definitely).

Memory and beliefs were then assessed in a phase immediately after corrections. Participants rated their belief in headlines and how well they remembered the veracity labels (fact-checks). Each headline from the initial phase appeared again without the corresponding images. Participants rated their belief from 1 (definitely false) to 10 (definitely true) and rated their memory for fact-checks from 1 (definitely corrected) to 10 (definitely affirmed). After seven days, participants repeated this procedure. Participants were then fully debriefed.

Figure 1

Schematic of the Procedure and Experimental Conditions



Note. **A**) Experiment timeline. **B)** Required ratings and example headlines in the false and true headline conditions across re-exposure and no re-exposure conditions.

Results

For the following results, we deviated from our preregistration and opted to report the outcomes from analyses using linear mixed-effect models (LMMs) instead of ANOVAs, because LMMs can account for correlations between repeated measures for each participant and inherent differences between headlines by including random effects. Results using the preregistered ANOVA-based approach (Supplementary Material Section C) differed from LMM-based results in that no significant effect of re-exposure on belief reduction was found. All other results were consistent between the two analysis types.

LMMs were run with the lme4 package (Bates et al., 2015). We used Satterthwaite's approximation to calculate p-values with the lmertest package (Kuznetsova et al., 2017) in R 4.2.1 (R Core Team, 2022). To make memory ratings for corrections more interpretable, we reversed-coded those ratings so that higher ratings indicated better memory for corrections. Results remained consistent after accounting for age and level of education and including participants who failed attention checks.

Re-exposure Reduces Belief in False Headlines and Improves Memory for Corrections

We examined potential differences in correction effects on belief change by testing for the effects of re-exposure and phase, as well as their interaction, on beliefs in false headlines (Table S2). Figure 2A shows that re-exposure predicted lower beliefs (b = -0.12, SE = 0.05, p = .020), and that beliefs did not change over the one-week delay (b = -0.02, SE = 0.05, p = .747). There was no significant interaction (b = -0.01, SE = 0.10, p = .911).

We next examined whether memory for corrections differed between correction conditions and over time (Table S3). Figure 2B shows that re-exposure predicted improved memory for corrections (b = 0.19, SE = 0.05, p < .001), and that the one-week delay reduced memory for corrections (*b* = -0.75, *SE* = 0.05, $p < .001$). There was no significant interaction (*b* = -0.13, *SE* = 0.10, *p* = .198).

Better Memory for Corrections is Associated with Reduced Belief in False Headlines

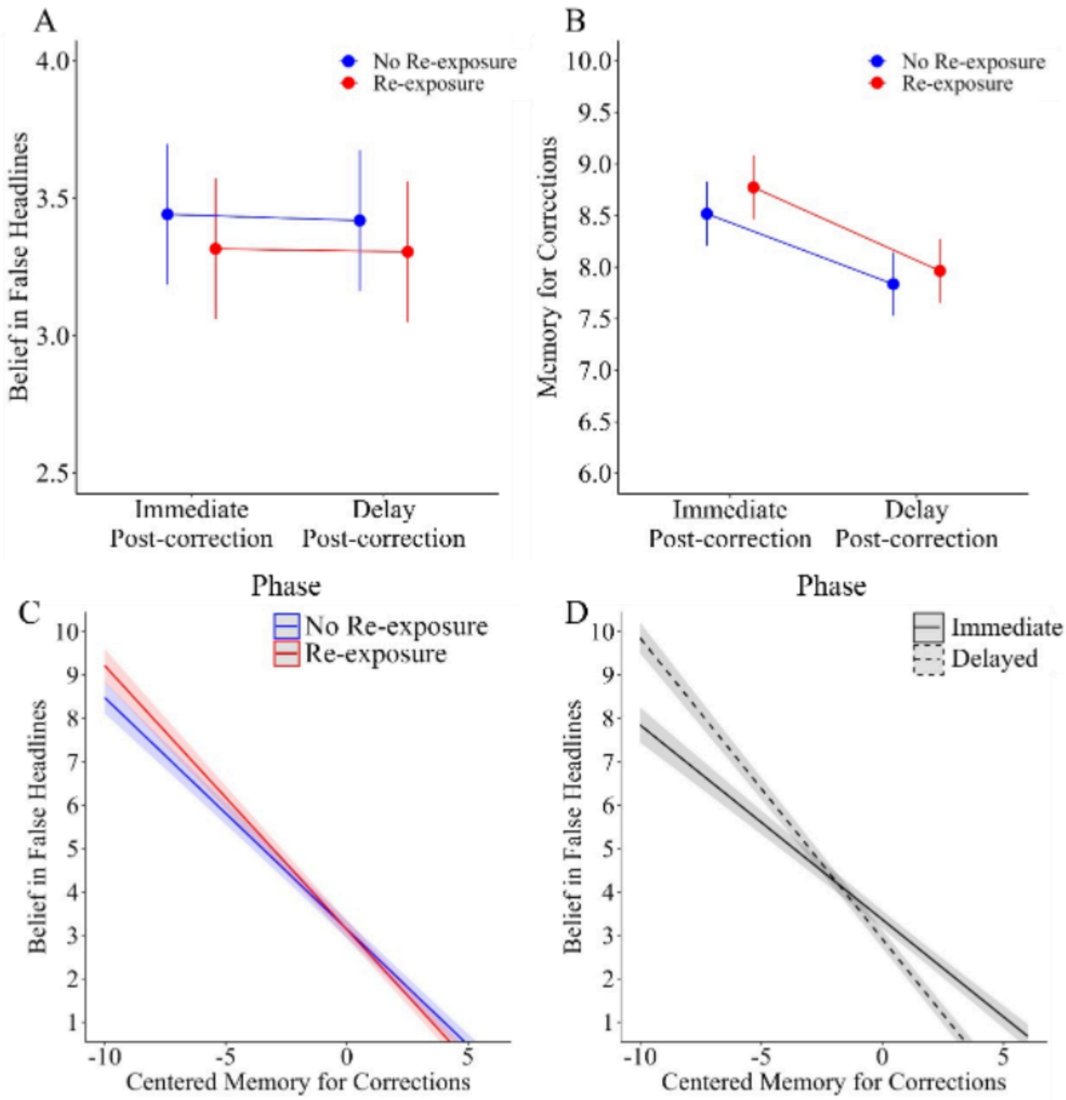
Next, we examined whether memory for corrections predicted belief in false headlines and whether this relationship varied across conditions and phases (Table S4). As hypothesized, Figure 2C shows that better memory for corrections was associated with lower false belief (*b* = -0.57, *SE* = 0.01, $p < .001$), and that this effect interacted with image re-exposure (*b* = -0.07, *SE* = 0.02, $p < .001$) such that the relationship between memory and belief was stronger in the re-exposure condition (*b* = -0.61, *SE* = 0.02, $p < .001$) than the no re-exposure condition (*b* = -0.53, *SE* = 0.02, $p < .001$). Figure 2D shows that this effect also interacted with phase (*b* = -0.25, *SE* = 0.02, $p < .001$), such that the relationship between memory and belief was stronger in the delayed (*b* = -0.69, *SE* = 0.02, $p < .001$) than immediate (*b* = -0.45, *SE* = 0.02, $p < .001$) post-correction phase.

Re-exposure Increases Belief in Corrections

We next examined whether re-exposure would engage the realism heuristic (Sundar, 2008), potentially decreasing belief in corrections. However, we found that re-exposure predicted increased belief in corrections (*b* = 0.18, *SE* = 0.07, *p* = .009, Table S5).

Figure 2

Belief and Memory for False Headlines



Note. Estimates from mixed effects models showing **A**) belief in false headlines and **B**) memory for corrections in re-exposure and no re-exposure conditions, in immediate and delayed phases. Regression lines showing relationship between centered memory for corrections and belief in false headlines across **C**) re-exposure conditions and **D**) phases. Shaded areas and error bars represent 95% confidence intervals.

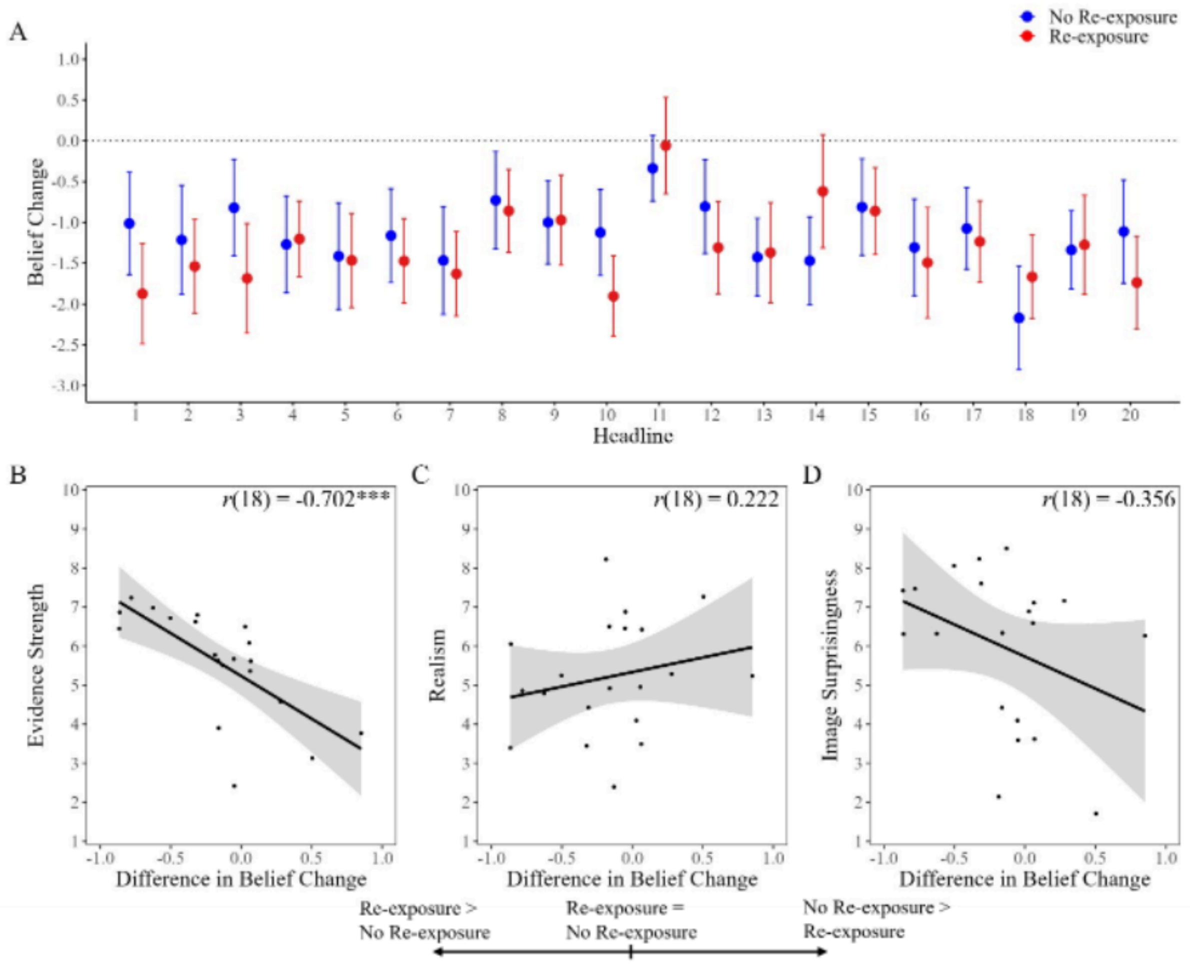
Motivation for Experiment 2 (Not Preregistered)

Although showing images alongside corrections reduced belief in false headlines to a greater extent than corrections without images, the effect was small (b = -0.12, indicating that re-exposure decreased belief by 0.12 points on a 10-point scale). This may have been because certain headlines were more amenable to re-exposure effects than others. To examine this possibility, we calculated the belief change between the initial and immediate phases for each headline stimulus, separating re-exposure and no re-exposure conditions. This difference score reflected how much beliefs were reduced, with more negative scores indicating more effective corrections. Figure 3A shows that re-exposure was nominally more effective for 13 headlines while no re-exposure was more effective for the other seven headlines.

We then compared difference scores across re-exposure conditions to estimate the difference in belief change for each headline. More negative scores indicate that re-exposure reduced beliefs more effectively, whereas more positive scores indicate that no re-exposure reduced beliefs more effectively. We examined whether this index was correlated with ratings of evidence strength, realism, and surprisingness for the headline images obtained from an earlier study (Guo, Zhong, et al., 2025). Figure 3B shows that evidence strength was negatively correlated with the difference in belief change, $r(18) = -0.702$, $p < .001$. Realism and surprisingness were not correlated with the difference in belief change (Figure 3C, D). Although correlational, this association could suggest that re-exposure to AIVM increased the efficacy of corrections for more probative images, whereas no re-exposure benefited less probative images.

Figure 3

Headline-level Belief Change and Correlations with Image Properties



Note. **A)** Difference between immediate post-correction and initial belief. **B)** Headline-level relationships between difference in belief change and evidence strength, **C)** realism, and **D)** image surprisingness. Negative values indicate that re-exposure reduced beliefs more, while positive values indicate that no re-exposure reduced beliefs more. Shaded regions and error bars are 95% confidence intervals.

Discussion

Experiment 1 showed that re-exposure to AIVM during corrections reduced beliefs in false headlines and enhanced memory for corrections. This suggests that re-presenting images during corrections facilitated associative encoding of correction headlines and the AIVM images, consistent with the picture superiority effect (Hockley & Bancroft, 2011). However, the effect of re-exposure on belief was small (b = -0.12), which may reflect heterogeneity among headlines. Indeed, post-hoc analyses found that the effectiveness of re-exposure differed for headlines with probative and non-probative images. Re-exposure increased belief in corrections, contrary to the prediction that realistic visual evidence would activate the realism heuristic (Messaris & Abraham, 2001; Sundar, 2008) and reduce belief in corrections. Even if the AIVM images seemed realistic, debunking them may have mitigated heuristic processing and engaged more analytic memory updating processes (e.g., the Heuristic-Systematic Model; Chaiken, 1980). Echoing prior research on the important role of memory in correcting misinformation (Kemp, Goldman, et al., 2024), memory for corrections predicted reduced beliefs in false headlines, and people relied more on memory for corrections after a delay.

Experiment 2

The primary goal of Experiment 2 was to directly manipulate the evidence strength of images to determine how it may interact with re-exposure during corrections. While there is a paucity of research on the relationship between probative images and false beliefs, semantically related word-picture pairs are often better remembered than unrelated pairs (Baadte & Meinhardt-Injac, 2019; Guillaume et al., 2017), suggesting that false headlines with probative AIVM could be remembered better. We therefore predicted that re-presenting probative AIVM images during belief correction would lead to better memories for corrections and thus stronger correction effects.

The secondary goal of Experiment 2 was to optimize the research methods. First, we increased statistical power based on effect sizes obtained in Experiment 1. Second, we preregistered our model specifications to reduce biased model reporting. Third, we removed belief ratings for corrections because they could imply that corrections could be inaccurate. Instead, we replaced them with surprise ratings, which were less likely to foster skepticism while still encouraging participants to attend to corrections.

Method

For full materials, power analysis, exclusion criteria, and procedure details, refer to Supplementary Material Section E.

Participants

After exclusions, the final sample included 368 participants (165 women, 198 men, and 5 unknown gender) from Prolific (M_{age} = 41.1, *SD* = 12.6, range = 18 to 79 years old) with the same inclusion criteria as Experiment 1. An a priori power analysis determined a target sample size of 450, and data collection continued until this target was reached. Ninety participants were excluded with the same exclusion criteria as Experiment 1.

Design

This experiment used a 2 (AIVM: re-exposure vs. no re-exposure) × 2 (Probity: probative vs. non-probative) × 2 (Headline veracity: true vs. false) within-participants design. Like Experiment 1, participants provided belief ratings in three phases (initial, immediate and delayed) and rated memory for corrections in immediate and delayed phases.

Materials

We conducted pretests to determine the strength of evidence that images provided for headlines as a basis for selecting images for the probative and non-probative conditions, and selected 32 false headline and 16 true headline sets for the experiment. Figure 4 shows examples of probative and non-probative images.

Procedure

The experiment timeline, delivery medium, and attention checks were identical to Experiment 1. For details on small changes made to randomization, refer to Supplementary Material Section E.

On the initial rating task, participants received the same instructions and made belief ratings as in Experiment 1. They viewed 16 true headlines paired with real images (eight with probative images, eight with non-probative images), and 32 false headlines (16 with probative images, 16 with non-probative images) in a randomized order. Next, on the correction/affirmation task, participants were told that headlines would be corrected or affirmed and appear with or without images, and that they were to rate their surprise to each correction or affirmation. These fact-checks appeared as in Experiment 1. Participants completed surprise ratings for each trial on a scale from 1 (not surprised at all) to 10 (extremely surprised). Finally, memory and beliefs were assessed immediately and one week after the correction phase, as in Experiment 1.

Figure 4

Example Probative and Non-Probative Images for False Headlines

Note. Example of the same false headline in the **A)** probative and **B)** non-probative condition.

Results

The pattern of results remained consistent after accounting for age and level of education, and after including participants who failed attention checks. We initially preregistered analyses including the delayed phase as exploratory because our sample size calculation aimed to detect the interaction between re-exposure and probity on only the immediate day 1 task. However, a sensitivity analysis indicated this sample size provided 97.50% [*95% CI* = 96.33, 98.38] power to detect the smallest effect size of interest which was the effect of probity on belief in false headlines (b = -0.11). Therefore, we present analyses including the delayed phase for consistency with Experiment 1, and include the preregistered analysis in Supplementary Material Section F. We report where results differ between this and the preregistered analysis.

Re-exposure and Probative Images Reduce Belief in False Headlines

We first examined how re-exposure conditions affected belief in false headlines (Table S6). Figure 5A shows that re-exposure significantly decreased belief in false headlines (b = -0.16, SE = 0.03, p < .001). Beliefs were lower for false headlines that were originally paired with probative images (b = -0.11, SE = 0.03, p < .001), and belief was significantly higher after a one-week delay (b = 0.26, SE = 0.03, p < .001). No interactions were significant (*p*s > .338). Moreover, modeling probity as a continuous fixed effect using pretest ratings ranging from 1 (weak evidence) to 10 (strong evidence) showed the same results.

Re-exposure and Probative Images Improve Memory for Corrections

We next examined the effect of re-exposure on memory for corrections (Table S7). Figure 5B shows that memory for corrections was significantly higher after re-exposure (b = 0.19, SE = 0.03, p < .001) and for false headlines with probative images (b = 0.11, SE = 0.03, p < .001). Memory for corrections also decreased significantly after a delay (b = -0.85, SE = 0.03, p < .001). There was a significant interaction between probity and phase (b = 0.21, SE = 0.05, p < .001), but no other significant interactions (*p*s > .252). Pairwise comparisons showed that memory for corrections in the immediate post-correction phase was not significantly different between the probative (*emmean* = 8.98, SE = 0.10) and non-probative (*emmean* = 8.97, SE = 0.10) conditions, z ratio = -0.13, p = .897. However, memory for corrections in the delayed post-correction phase was significantly greater in the probative (*emmean* = 8.24, SE = 0.10) than non-probative (*emmean* = 8.02, SE = 0.10) condition, z ratio = -5.83, p < .001. These results showed that memory for corrections was selectively enhanced for probative images after a delay, regardless of correction format.

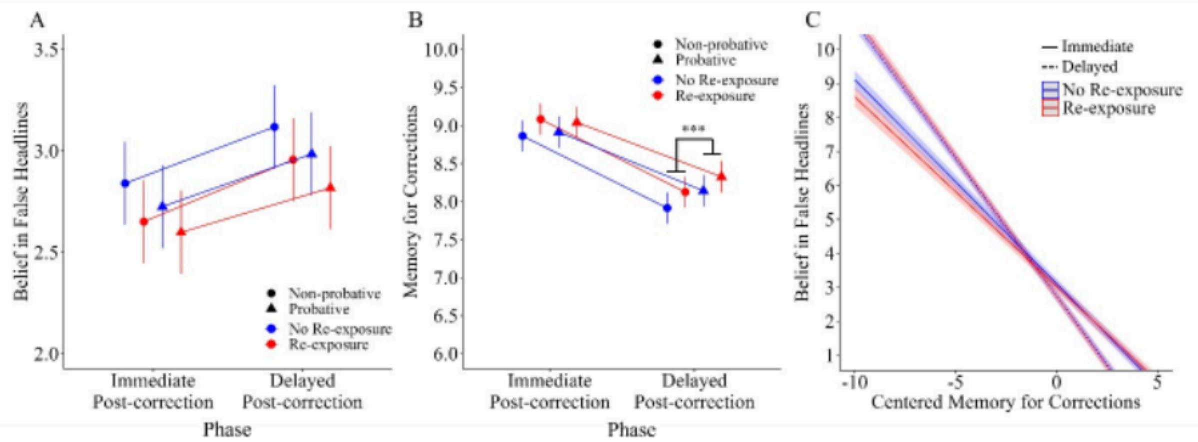
It is important to note that in the preregistered analyses with only the immediate phase, probity was only a significant predictor of belief when values were z-scored (Table S8), and probity was not a predictor of memory.

Better Memory for Corrections is Associated with Reduced Belief in False Headlines

We were also interested in whether the relationship between memory and belief changed after a delay, and excluded the probity fixed effect to replicate our approach in Experiment 1 (Table S9). Figure 5C shows that stronger memory for correction ratings was associated with lower belief in false headlines (b = -0.70, SE = 0.01, p < .001); a significant memory × phase interaction (b = -0.24, SE = 0.01, p < .001) showed that this relationship was stronger in the delayed (b = -0.82, SE = 0.01, p < .001) than immediate post-correction phase (b = -0.58, SE = 0.01, p < .001), z ratio = 22.09, p < .001. Extending Experiment 1, a significant three-way interaction (b = -0.06, SE = 0.02, p = .003) qualified these effects. Pairwise comparisons showed that in the immediate phase, the negative association was significantly greater after no re-exposure (b = -0.60, SE = 0.01) than re-exposure (b = -0.56, SE = 0.01), z ratio = -2.68, p = .009. Conversely, in the delayed phase, the association was not significantly different between no re-exposure (b = -0.81, SE = 0.01) and re-exposure (b = -0.82, SE = 0.01), z ratio = 1.36, p = .173.

Figure 5

Belief in False Headlines and Memory for Corrections on Immediate and Delayed Tests



Note. Estimates from models showing **A**) belief in false headlines and **B**) memory for corrections in the re-exposure, no re-exposure, probative and non-probative conditions, across immediate and delayed post-correction phases. **C**) Regression lines showing relationship between centered memory for corrections and belief in false headlines across phase and re-exposure conditions. Error bars and shaded areas are 95% confidence intervals.

Discussion

Experiment 2 tested whether the effects in Experiment 1 would replicate with a larger sample and whether image probity moderates re-exposure effects. Replicating Experiment 1, re-exposure reduced belief in false headlines and improved memory for corrections, and stronger memory for corrections was associated with lower false belief. These effects emerged for probative and non-probative images, suggesting that repeating images during corrections can benefit belief change regardless of evidence strength. Contrary to our hypothesis, probative images did not selectively enhance re-exposure effects on false belief reduction. However, false headlines paired with probative AIVM were associated with greater belief reduction and better memory for corrections overall. One possibility is that probative images were more persistently associated with headlines because they were more semantically related (see Supplementary Material Section H), such that retrieving those images benefited memory for corrections and reduced false belief regardless of re-exposure.

Experiment 3

Experiments 1 and 2 showed that re-exposure to AIVM during corrections reduced false beliefs and improved memory for corrections. However, corrections in the re-exposure condition mentioned AI-generated imagery, raising the possibility that the effects reflected references to AI rather than image re-exposure per se. Experiment 3 addressed this issue by using identical correction texts across conditions. Because the primary goal was to rule out this explanation, Experiment 3 assessed belief and memory immediately after correction and did not include a delayed test. If the immediate test effects replicated those observed in Experiment 2, the delayed pattern would be expected to replicate that experiment as well.

Method

For full materials, power analysis, exclusion criteria, and procedure details, refer to Supplementary Material Section I.

Participants

After exclusions, the final sample included 293 participants (157 women, 134 men, and 2 unknown gender) from Prolific (M_{age} = 44.2, *SD* = 14.2, range = 18 to 80 years old) with the same inclusion criteria as Experiments 1 and 2. An a priori power analysis determined a target sample size of 300, and data collection continued until this target was reached. Ten participants were excluded with the same criteria as Experiments 1 and 2.

Design

This experiment used a 2 (AIVM: re-exposure vs. no re-exposure) × 2 (Probity: probative vs. non-probative) × 2 (Headline veracity: true vs. false) within-participants design. Participants provided belief ratings in initial and immediate post-correction phases, and memory ratings in the immediate post-correction phase.

Materials

Since Experiment 3 took place nine months after Experiment 2, we reassessed the veracity of the headlines before selecting items for Experiment 3 and replaced three false headlines.

Procedure

The delivery medium, attention checks, measures, and randomization were identical to Experiment 2, but Experiment 3 did not include a delay phase. The major change in Experiment 3 was that re-exposure and no re-exposure corrections and affirmations did not mention AI (i.e., corrections read “this is FALSE”, and affirmations read “this is TRUE”). Furthermore, participants were encouraged to use the full scale to make precise ratings.

Results

The patterns of results remained consistent after accounting for age and level of education and including participants who failed attention checks. For additional preregistered analyses on true headlines, see Supplementary Material Section J.

Preregistered Analyses

Re-exposure and Probative Images Reduce Belief in False Headlines

We first examined whether the relationship between re-exposure and belief was consistent with Experiments 1 and 2 after removing the mention of AI (Table S10). Figure 6A shows that Experiments 1 and 2 were replicated: re-exposure predicted reduced belief in false headlines (b = -0.24, SE = 0.04, p < .001), and there was no significant interaction (b = 0.10, SE = 0.08, p = .225). Like Experiment 2, probative images were associated with reduced false belief (b = -0.11, SE = 0.04, p = .007). Analyses with continuous probative scores showed the same results.

Re-exposure and Probative Images Improve Memory for Corrections

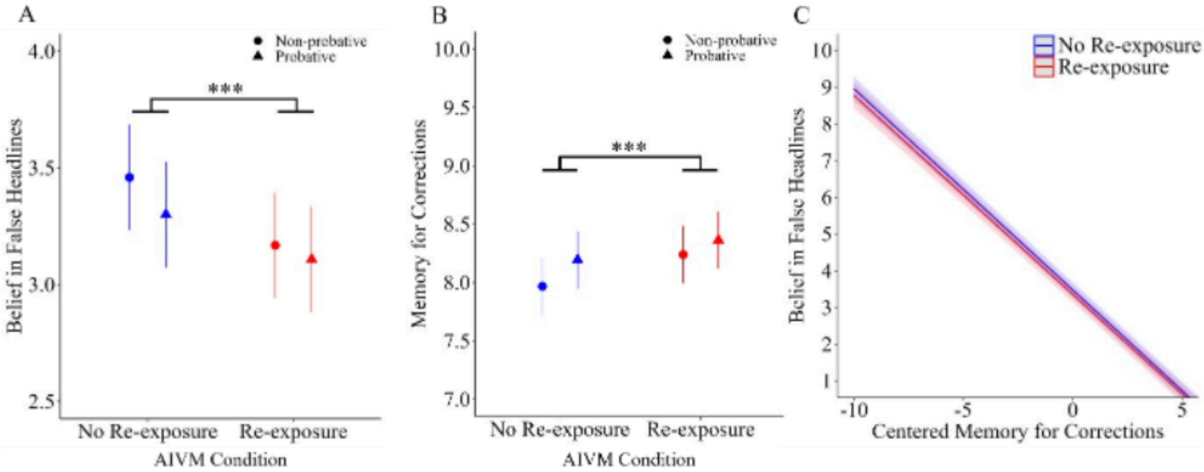
We next examined memory for corrections (Table S11). Figure 6B shows that results replicated Experiments 1 and 2: re-exposure predicted improved memory for corrections (b = 0.22, SE = 0.04, p < .001), and there was no significant interaction (b = -0.10, SE = 0.08, p = .188). Like Experiment 2, probative images predicted better memory for corrections (b = 0.17, SE = 0.04, p < .001).

Better Memory for Corrections is Associated with Reduced Belief in False Headlines

Next, we examined the relationship between memory for corrections and belief in false headlines (Table S12). Consistent with Experiments 1 and 2, Figure 6C shows that higher memory ratings predicted lower false headline belief (b = -0.54, SE = 0.01, p < .001). In contrast to Experiments 1 and 2, there was no evidence for an interaction between memory and belief, b = 0.01, SE = 0.02, p = .757.

Figure 6

Belief in Headlines and Memory for Corrections



Note. Estimates from mixed effects models showing **A**) belief in false headlines and **B**) memory for corrections in the immediate post-correction phase, in the re-exposure, no re-exposure, probative and non-probative conditions. **C)** Regression lines showing relationship between centered memory for corrections and belief in false headlines. Error bars and shaded areas are 95% confidence intervals. ***$p < .001$.

Discussion

Experiment 3 tested whether re-exposure effects persist when correction wording is held constant across conditions. Replicating Experiments 1 and 2, re-exposure reduced belief in false headlines and improved memory for corrections, and better memory for corrections predicted lower belief in false headlines. These results suggest that the re-exposure effect from the prior experiments does not arise from informing participants about AI and instead reflects beneficial mnemonic effects of imagery, consistent with the picture superiority effect.

General Discussion

With the proliferation of AI-generated imagery, people are increasingly exposed to false claims accompanied by compelling visual evidence (Corsi et al., 2024). Fact-checkers attempt to correct such content but differ in whether they repeat the original images when issuing corrections. Across three experiments, we found that re-exposing AI-generated visual misinformation (AIVM) during corrections improved memory for corrections and reduced belief in false headlines immediately and after a one-week delay. These results are consistent with the picture superiority effect (Hockley, 2008; Hockley & Bancroft, 2011), suggesting that images facilitate associative encoding between headlines and corrections. In contrast, predictions derived from the realism heuristic (Messaris & Abraham, 2001; Sundar, 2008) were not supported: re-exposure enhanced correction effectiveness across both probative and non-probative images rather than increasing belief in misinformation. Probative images also improved memory for corrections and reduced belief regardless of re-exposure, possibly because their stronger semantic relation to headlines made them easier to retrieve and update during correction.

Conceptually replicating prior findings, memory for corrections was associated with lower belief in false headlines (Kemp, Sinclair, et al., 2024; Wahlheim et al., 2020), and this relationship strengthened after a delay (Swire-Thompson et al., 2023). In the immediate post-correction phase, memory for corrections could be generally strong across all headlines, and beliefs may be adjusted based on automatic fluency or familiarity processes. After a delay, some corrections are consolidated while others are forgotten (Guo et al., 2026), resulting in greater reliance on memory. However, the effect of re-exposure on this relationship was inconsistent across experiments, which may be due to unknown differences between samples.

There were limitations. First, participants saw an equal number of true and false headlines in Experiment 1 but viewed more false than true headlines in Experiment 2 and 3, which could have created a response bias toward judging headlines as false when uncertain (Altay et al., 2024). However, this may be more generalizable to fact-checking websites that heavily feature corrections. Second, we did not examine any topics with strong prior beliefs, which limits generalizability. Finally, corrections were simple labels, so it is unclear whether results will generalize to more detailed corrections.

Conclusion

The proliferation of AIVM highlights the need to identify effective ways to issue corrections. We found that re-exposure to AIVM during corrections reduced belief in false headlines and improved memory for corrections relative to corrections without images. These effects emerged even when the images' AI origins were not disclosed, generalized across probative and non-probative images, and persisted after a one-week delay. Although prior work shows that images can increase the credibility of misinformation, our findings indicate that when used within corrections, images can enhance their effectiveness. These results suggest that attaching corrections directly to the original AIVM may be an effective and scalable strategy for combating visual misinformation in everyday online contexts.

Authors' Contributions

SG: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Resources, Software, Visualization, Writing – original draft, Writing – review and editing. **SM:** Data curation, Formal Analysis, Validation, Writing – review and editing. **CW:** Conceptualization, Methodology, Supervision, Writing – review and editing. **XH:** Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Writing – review and editing.

References

Altay, S., De Angelis, A., & Hoes, E. (2024). Media literacy tips promoting reliable news improve discernment and enhance trust in traditional media. *Communications Psychology*, *2*(1). https://doi.org/10.1038/s44271-024-00121-5

Baadte, C., & Meinhardt-Injac, B. (2019). The picture superiority effect in associative memory: A developmental study. *British Journal of Developmental Psychology*, *37*(3), 382–395. https://doi.org/10.1111/bjdp.12280

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of personality and social psychology*, *39*(5), 752.

Corsi, G., Marino, B., & Wong, W. (2024). The spread of synthetic media on X. *Harvard Kennedy School Misinformation Review*, *5*(3), 1-19.

Diakopoulos, N., & Johnson, D. (2021). Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New Media and Society*, *23*(7), 2072–2098. https://doi.org/10.1177/1461444820925811

Ecker, U. K. H., Lewandowsky, S., Swire, B., & Chang, D. (2011). Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction. *Psychonomic Bulletin and Review*, *18*(3), 570–578. https://doi.org/10.3758/s13423-011-0065-1

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, *1*(1), 13–29. https://doi.org/10.1038/s44159-021-00006-y

Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the Unbelievable: Some Problems in the Rejection of False Information. *Journal of Personality and Social Psychology*, *59*(4), 601–613. https://doi.org/10.1037/0022-3514.59.4.601

Guillaume, F., Baier, S., Bourgeois, M., & Tinard, S. (2017). Format change and semantic relatedness effects on the ERP correlates of recognition: old pairs, new pairs, different stories. *Experimental Brain Research*, *235*(4), 1007–1019. https://doi.org/10.1007/s00221-016-4859-2

Guo, S., Chen, D., Hu, W., & Hu, X. (2026). Faster Initial Retrieval of Misinformation Corrections Predicts Better Long-Term Memory: An ERP Study. *Neuropsychologia*, 109428.

Guo, S., Swire-Thompson, B., & Hu, X. (2025). Specific media literacy tips improve AI-generated visual misinformation discernment. *Cognitive Research: Principles and Implications*, *10*(1). https://doi.org/10.1186/s41235-025-00648-z

Guo, S., Zhong, Y., & Hu, X. (2025). People are more susceptible to misinformation with realistic AI-synthesized images that provide strong evidence to headlines. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-189

Heley, K., Gaysynsky, A., & King, A. J. (2022). Missing the Bigger Picture: The Need for More Research on Visual Health Misinformation. In *Science Communication* (Vol. 44, Issue 4, pp. 514–527). SAGE Publications Inc. https://doi.org/10.1177/10755470221113833

Hockley, W. E. (2008). The picture superiority effect in associative recognition. *Memory and Cognition*, *36*(7), 1351–1359. https://doi.org/10.3758/MC.36.7.1351

Hockley, W. E., & Bancroft, T. (2011). Extensions of the Picture Superiority Effect in Associative Recognition. *Canadian Journal of Experimental Psychology*, *65*(4), 236–244.

Kemp, P. L., Goldman, A. C., & Wahlheim, C. N. (2024). On the role of memory in misinformation corrections: Repeated exposure, correction durability, and source credibility. In *Current Opinion in Psychology* (Vol. 56). Elsevier B.V. https://doi.org/10.1016/j.copsyc.2023.101783

Kemp, P. L., Loaiza, V. M., & Wahlheim, C. N. (2022). Fake news reminders and veracity labels differentially benefit memory and belief accuracy for news headlines. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-25649-6

Kemp, P. L., Sinclair, A. H., Adcock, R. A., & Wahlheim, C. N. (2024). Memory and belief updating following complete and partial reminders of fake news. *Cognitive Research: Principles and Implications*, *9*(1). https://doi.org/10.1186/s41235-024-00546-w

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/JSS.V082.I13

Maertens, R., Roozenbeek, J., Simons, J. S., Lewandowsky, S., Maturo, V., Goldberg, B., Xu, R. & van der Linden, S. (2025). Psychological booster shots targeting memory increase long-term resistance against misinformation. *Nature Communications*, *16*(1), 2062.

Messaris, P., & Abraham, L. (2001). The Role of Images in Framing News Stories. In *Framing Public Life: Perspectives on Media and Our Understanding of the Social World* (pp. 215–226). Routledge.

Midjourney (2025). Available from https://www.midjourney.com/home

Newman, E. J., Garry, M., Unkelbach, C., Bernstein, D. M., Lindsay, D. S., & Nash, R. A. (2015). Truthiness and falsiness of trivia claims depend on judgmental contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(5), 1337.

O'Rear, A. E., & Radvansky, G. A. (2020). Failure to accept retractions: A contribution to the continued influence effect. *Memory and Cognition*, *48*(1), 127–144. https://doi.org/10.3758/s13421-019-00967-9

Paivio, A. (1991). Dual Coding Theory: Retrospect and Current Status. *Canadian Journal of Psychology*, *45*(3), 255–287.

Politifact. (2026). https://www.politifact.com/

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Seo, K. (2020). Meta-Analysis on Visual Persuasion– Does Adding Images to Texts Influence Persuasion? *Athens Journal Of Mass Media And Communications*, *6*(3), 177–190. https://doi.org/10.30958/ajmmc.6-3-3

Snopes. (2026). https://www.snopes.com/

Sundar, S. S. (2008). The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility. *Digital Media, Youth, and Credibility*, 73–100. https://doi.org/10.1162/dmal.9780262562324.073

Sundar, S. S., Molina, M. D., & Cho, E. (2021). Seeing Is Believing: Is Video Modality More Powerful in Spreading Fake News via Online Messaging Apps? *Journal of Computer-Mediated Communication*, *26*(6), 301–319. https://doi.org/10.1093/jcmc/zmab010

Swire-Thompson, B., Dobbs, M., Thomas, A., & DeGutis, J. (2023). Memory failure predicts belief regression after the correction of misinformation. *Cognition*, *230*. https://doi.org/10.1016/j.cognition.2022.105276

Wahlheim, C. N., Alexander, T. R., & Peske, C. D. (2020). Reminders of Everyday Misinformation Statements Can Enhance Memory for and Beliefs in Corrections of Those Statements in the Short Term. *Psychological Science*, *31*(10), 1325–1339. https://doi.org/10.1177/0956797620952797

Wahlheim, C. N., Kemp, P. L., Wiley, R. W., & Engelhardt, A. M. (2025). Sustaining corrected beliefs in false news headlines over time: The roles of correction format and recognizing corrections. *Memory & Cognition*, 1-25. https://doi.org/10.3758/s13421-025-01760-7

Wellons, B. M., & Wahlheim, C. N. (2025). Misinformation reminders enhance belief updating and memory for corrections: The role of attention during encoding revealed by eye tracking. *Cognitive Research: Principles and Implications*, 10(1), 39. https://doi.org/10.1186/s41235-025-00649-y

Whitehead, P. S., Davis, C. P., Park, J. S., & Marsh, E. J. (2025). Pictures are not always worth a thousand words: Nonprobative pictures did not increase the effectiveness of misinformation corrections. *Journal of Applied Research in Memory and Cognition*. https://doi.org/10.1037/mac0000214

Wijnker, W., Smeets, I., Burger, P., & Willems, S. (2022). Debunking strategies for misleading bar charts. *Journal of Science Communication*, *21*(7), A07.