HHS Public Access

Author manuscript

Cognition. Author manuscript; available in PMC 2023 December 10.

Published in final edited form as:

Cognition. 2023 January ; 230: 105276. doi:10.1016/j.cognition.2022.105276.

Memory failure predicts belief regression after the correction of misinformation

Briony Swire-Thompson[a,b,*], **Mitch Dobbs**[a], **Ayanna Thomas**[c], **Joseph DeGutis**[d,e]

[a]Network Science Institute, Northeastern University, Boston, USA

[b]Institute of Quantitative Social Science, Harvard University, Cambridge, USA

[c]Department of Psychology, Tufts University, Cambridge, USA

[d]Boston Attention and Learning Laboratory, VA Boston Healthcare System, Boston, MA, USA

[e]Department of Psychiatry, Harvard Medical School, Boston, MA, USA

Abstract

After misinformation has been corrected, people initially update their belief extremely well. However, this change is rarely sustained over time, with belief returning towards pre-correction levels. This is called belief regression. The current study aimed to examine the association between memory for the correction and belief regression, and whether corrected misinformation suffers from belief regression more than affirmed facts. Participants from Prolific Academic (N = 612) rated the veracity of 16 misinformation and 16 factual items and were randomly assigned to a correction condition or test-retest control. Immediately after misinformation was corrected and facts affirmed, participants re-rated their belief and were asked whether they could remember the items' presented veracity. Participants repeated this post-test one month later. We found that belief and memory were highly associated, both immediately (ρ = 0.51), and after one month (ρ = 0.82), and that memory explained 66% of the variance in belief regression after correcting for measurement reliability. We found the rate of dissenting (accurately remembering that misinformation was presented as false but still believing it) remained stable between the immediate and delayed post-test, while the rate of forgetting quadrupled. After one month, 57% of participants who believed in the misinformation thought that the items were presented to them as true. Belief regression was more pronounced for misinformation than facts, but this was greatly attenuated once pre-test belief was equated. Together, these results clearly indicate that memory plays a fundamental role in belief regression, and that repeated corrections could be an effective method to counteract this phenomenon.

*Corresponding author at: Network Science Institute, Northeasetern University, 177 Huntington Avenue, Boston, USA. b.swire-thompson@northeastern.edu (B. Swire-Thompson).

Author contributions

BST was responsible for study conceptualization and participant recruitment. BST, JD, and AT designed the study. BST and MD conducted the data analysis, BST drafted the manuscript and JD, MD, and AT provided critical revisions. All authors approved the final version of the manuscript for submission.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2022.105276.

Keywords

Misinformation; Belief regression; Memory; Belief updating

1. Introduction

After misinformation has been corrected, individuals initially update their belief extremely well. However, this belief change is rarely sustained over time, with participants' belief in misinformation returning towards their pre-correction levels (Kowalski & Taylor, 2017; Swire, Ecker, & Lewandowsky, 2017). We refer to the phenomenon where people appear to re-endorse or "re-believe" in the original misinformation over time as *belief regression*. Where the *continued influence effect* refers to the general continued use of corrected misinformation in memory and reasoning, belief regression is the temporal impermanence of the correction's efficacy. The vast majority of misinformation research examines belief change immediately after corrections have been presented (Dias & Sippitt, 2020), despite longer-term belief change being more relevant to the real world. Studies that include a delayed retention interval have found belief regression to be a robust phenomenon (Berinsky, 2017; Carey et al., 2022; Rich & Zaragoza, 2020), although its mechanisms remain unknown. We aimed to examine the association between memory and belief regression after the correction of misinformation, and whether corrected misinformation suffers more from belief regression than affirmed facts.

1.1. Memory and belief updating

It has long been established that memory plays a vital role in the correction of misinformation (Seifert, 2002), yet the exact association between memory and belief remains unknown. Intuitively, in order for a person to believe an item to be false, they must remember that it was presented as false. Indeed, Wahlheim, Alexander, and Peske (2020) found that belief in misinformation was lower when corrections were remembered than when they were not, and that corrective reminders increased the accuracy of participants' beliefs. The cognitive mechanisms usually assumed to be motivating the continued influence effect often rely upon models of memory—namely that there have been failures to either retrieve the correct information, or failures to integrate the new information into one's mental model (see Sanderson & Ecker, 2020). The mental-model account assumes that people create "mental models" of events or causal situations, and the continued influence effect occurs when a relevant correction is encountered, yet there is a failure to integrate and update the model. Studies from the educational literature have also relied upon memory models to explain the success of *refutational corrections*, where the misconception is directly followed by evidence explaining why it is incorrect. Kendeou and O'Brien (2014) argue that it is the co-activation of the misconception and corrective information that facilitates the new information's integration into memory representations.

Belief in the corrective information (and thus disbelief in the misinformation) is also integral to successful belief updating. To illustrate, O'Rear and Radvansky (2020) investigated whether the continued influence effect was partially due to people failing to believe corrections. Participants read a fictitious scenario involving a minibus accident, where

passengers were first presented as elderly, and subsequently revealed to *not* be elderly. A large proportion of participants, over 40% of those who remembered the correction, did not believe in the veracity of the correction. The participants who did not accept the correction as valid used the misinformation in their inferential reasoning at a similar rate as those who never received a correction at all. In other words, participants continued to use the outdated misinformation in their reasoning simply because they did not believe that the correction was accurate or genuine.

1.2. Memory and belief regression

A core aspect of paradigms that include a delay is that participants must recall the information as false to maintain belief change over time.[1] Several findings suggest that particular aspects of memory failures may underlie belief regression. Gilbert (1991) proposed that understanding information as false is more effortful than the acceptance of information as true, and it is often assumed that a "false tag" is attached to inaccurate statements as a contextual detail. Failures of recollection, a process thought to allow for the retrieval of contextual details (Yonelinas, 2002), could lead to the retrieval of the information without the false tag, resulting in its inadvertent acceptance (Schacter, 2008). Older adults are more prone to belief regression than younger adults (Swire, Ecker, & Lewandowsky, 2017), which aligns well with age-related memory decline and older adults being particularly poor at recollection (Dennis, Gutchess, & Thomas, 2020).

However, it is possible that people on occasion correctly recall the false tag, but increase belief in the corrected misinformation nonetheless. For instance, people could forget other associated details such as the specific reasons explaining *why* misinformation is false, or that the information came from a reputable source. Conceptually, the latter would be the opposite of the *sleeper effect* (Hovland, Lumsdaine, & Sheffield, 1949), where messages become more persuasive over time due to the forgetting or dissociation of the message and the disreputable source. In sum, it is important that the correction remains persuasive as time passes (Hill, Lo, Vavreck, & Zaller, 2013). Rich, Van Loon, Dunlosky, and Zaragoza (2017) found participants who believed in corrective feedback updated their belief more frequently, and were also more likely to report correct answers one week later than those who did not. Thus, belief regression may partially occur due to participants increasingly "dissenting": accurately remembering that the misinformation was presented as false but maintaining belief in the misinformation.

1.3. Asymmetry in belief regression

Finally, it is still an open question whether belief regression is asymmetrical, with affirmed facts showing more sustained belief change than corrected misinformation (Skurnik et al. 2007, as cited by Schwarz, Sanna, Skurnik, & Yoon, 2007). Importantly, asymmetry would suggest that different mechanisms underlie fact affirmation and misinformation correction (see Swire, Berinsky, Lewandowsky, & Ecker, 2017). However, another potential reason for this asymmetry is that misinformation and fact stimuli used in research already have

[1]This is in contrast to directed forgetting, where participants are instructed to forget presented items (Zacks, Radvansky, & Hasher, 1996), or the recall to reject paradigm, where participants recall presented items to reject new foils (Gallo, Bell, Beier, & Schacter, 2006).

asymmetrical beliefs prior to affirmations/corrections. In other words, facts might already be perceived as more true than misinformation is perceived to be false. This would make it more difficult to correct misinformation, given that further belief change is necessary. It would also create unequal scaling issues such as regression to the mean, given that the facts are closer to the ceiling than misinformation is to the floor. Previous literature reporting asymmetric belief updating has either not measured pre-existing beliefs at all (Skurnik et al. 2007 as cited by Schwarz et al. (2007); Peter & Koch, 2016), or has not controlled for it (Swire, Ecker, & Lewandowsky, 2017). We aim to determine if belief regression asymmetry still occurs after correcting for scaling effects, which could potentially provide support for mechanistic differences between misinformation and facts.

1.4. The current study

The goals of this study were to investigate (1) the association between memory for the correction and belief in the misinformation both immediately and after a one-month delay; (2) the amount of variance that memory accounts for in belief regression; (3) the degree to which belief regression occurs due to "dissenting" (accurately remembering that the misinformation was presented as false but still believing in it) or forgetting (inaccurately remembering that the misinformation was presented as true and believing in it), and (4) to replicate whether corrected misinformation suffers from more belief regression than affirmed facts (i.e., belief regression asymmetry), even after items are equated at pre-test.

2 Methods

2.1. Design

This was a longitudinal study with three groups (correction memory-first vs. correction belief-first vs. test-retest control) assessed at three different time points (pre-test, immediate post-test, and one-month delayed post-test). We counterbalanced post-test memory and post-test belief blocks to examine if participants changed their belief ratings based upon prior memory ratings (or vice versa).

2.2. Participants

Prolific Academic was selected to recruit participants, as it is more diverse in age, race, and education than other samples of convenience (Peer, Brandimarte, Samat, & Acquisti, 2017). The only selection criteria were that the participants were over the age of 18 and from the U.S.A. There were 699 participants recruited in the pre-test and 612 completed the one-month post-test (88% retention). In our final sample there were 298 males, 298 females, and 16 individuals choosing not to disclose their gender. Participants' age ranged between 18 and 75 (M = 33.78, SD = 11.86, see Fig. S1). Participants were randomly assigned to either the correction with memory first condition (N = 202), correction with belief first condition (N = 208), or the control condition (N = 202; see Table S1 for demographic distributions across conditions).

2.3. Procedure

Using Qualtrics software (Qualtrics, Provo, UT), participants first read a Northeastern University approved consent form (#19–04–90) and agreed to participate in the study. In

the pre-test, all participants rated 16 facts and 16 misinformation items in a randomized order for (a) how much they believed them to be true (0 = definitely false; 5 = unsure, and 10 = definitely true), and (b) how much they had considered the claim in the past (0 = not at all; 10 = a great deal). They were instructed to answer higher on the scale if they had spent a long time contemplating, considering, or deliberating the information. Items were presented on the screen one at a time and all items were rated before moving on. Participants in the correction conditions were next shown corrections (for the misinformation) and affirmations (for the facts), which also appeared on the screen one at a time in a randomized order. For each item, participants were asked to rate how surprised they were on a 0–10 scale to ensure that they read the corrections and affirmations (as in Swire-Thompson, Miklaucic, Wihbey, Lazer, & DeGutis, 2022).

In the immediate post-test, all participants re-rated their beliefs for each item, presented in a randomized order. The instructions were “Next, please rate the same statements again on a 0–10 scale for whether or not you **believe** them to be true” (0 = definitely false; 5 = unsure, and 10 = definitely true). Participants in the correction conditions also completed a memory block, where they were told “This is now a memory test. Can you **remember** whether we told you that these statements were true or false?” Participants were asked to respond on a 0–10 scale (0 = definitely false; 5 = unsure, and 10 = definitely true), and all items were presented on one page in a randomized order. The memory block and belief block were counterbalanced such that half of the correction condition participants received the memory block first and half received the belief block first. After one month, participants were invited to participate in the delayed post-test, which was identical to the immediate post-test. Participants in the control condition rated their belief again, while those in the correction conditions re-rated both memory and belief in their assigned order.

2.4. Stimuli

Stimuli were 32 items selected primarily from Swire-Thompson et al. (2022) chosen for their high test-retest reliability (see Swire-Thompson et al., 2022 for a description of how they were created). Selected items ranged from $\rho = 0.50$ to $\rho = 0.75$, and all corrections were designed to have similar word counts ($M = 57.44$, $SD = 3.60$). See Table 1 for an example item, Table S2 for all misinformation, and Table S3 for all facts. All corrections repeated the initial statement, included a false tag, a reputable source, and provided an explanation as to why the misinformation was false. Similarly, all affirmations repeated the initial statement, included a true tag, a reputable source, and provided an explanation as to why it was true. Participants’ pre-test belief rating was also displayed at the bottom of each correction and affirmation (as in Swire-Thompson et al., 2022) to promote co-activation between original belief and the correction or affirmation. The overall test-retest reliability of the misinformation stimuli (aggregated across all items, calculated from the control group) was $\rho = 0.93$ and $\rho = 0.80$ for the immediate and delayed conditions, respectively. The test-retest reliability of the fact stimuli was $\rho = 0.92$ and 0.71 for the immediate and delayed conditions, respectively.

2.5. Sample size justification

Belief regression effect sizes vary substantially ranging from $\eta p^2 = 0.04$ (small) to 0.43 (large; Rich & Zaragoza, 2020). One main analysis was a 2 × 2 repeated measures ANOVA, with the dependent variable misinformation belief, within-subjects factor post-test retention interval (immediate vs. delayed), and between-subjects factors correction (correction vs. control). If we take the smallest effect size from previous studies ($\eta p^2 = 0.04$), a power analysis conducted by G*Power3 (Faul, Erdfelder, Buchner, & Lang, 2009) recommends a sample size of 80 participants (for effect size $f = 0.20$ with $\alpha = 0.05$, power = 0.95 and a moderate correlation between repeated measures, $r = 0.50$). Given that the memory component of this study is unknown, particularly with regards to how memory and belief interact, we aimed to boost the sample size substantially to detect an effect of $f = 0.10$, requiring 328 participants. We further boosted the total sample to a total of >600 to achieve 400 participants in the correction conditions. This provided added sensitivity to detect individual differences associated with the belief regression index.

3. Results

3.1. Participants

We first sought to determine if our correction and control groups were well-balanced across demographic and baseline belief variables. There were no significant differences between the correction memory-first group, correction belief first-group, and control group for age ($p = .805$), education ($p = .534$), gender ($p = .249$), partisanship ($p = .453$), or baseline beliefs ($p = .129$). See Table S1 for details.

3.2. The influence of reporting memory prior to belief

We next investigated whether rating memory prior to belief influenced belief ratings. Focusing on the correction conditions only (i.e., disregarding the control), a 2 × 2 between-within ANOVA with factors block order (memory first vs. belief first) and retention interval (immediate post-test, delayed post-test) on belief ratings did not reveal any main effects of block order ($p = .251$) nor an interaction with retention interval ($p = .519$). We replicated this analysis using Bayes factors (BFs) given that this method can quantify relative evidence favoring the null hypothesis. The findings can be expressed as either BF_{10} which quantifies support for the alternative hypothesis, or BF_{01} which quantifies support for the null hypothesis. A BF between 1 and 3 provides anecdotal evidence, 3–10 provides moderate evidence, 10–30 provides strong evidence, 30–100 provides very strong evidence, and a BF >100 constitutes extreme evidence (Wagenmakers, Marsman, Jamil, et al., 2018). We found moderate evidence that there was no main effect for block order ($BF_{01} = 4.99$) nor block order × retention interval interaction ($BF_{01} = 6.76$).

For completeness, we also tested whether rating belief first influenced subsequent memory ratings. We conducted a 2 × 2 ANOVA with factors block order (memory first vs. belief first), retention interval (immediate post-test vs. delayed post-test) on memory ratings. There was no main effect of block order ($p = .367$), nor a retention interval × block order interaction ($p = .778$). We also found moderate evidence using BFs that there was no main effect for block order ($BF_{01} = 7.50$) nor block order × retention interval interaction ($BF_{01} =$

9.89). Given that presentation order did not impact post-test memory or belief, we collapsed across block order in all subsequent analyses.

3.3. The association between memory and belief after corrective information

In order to investigate the association between memory for the correction and belief in the misinformation, we ran correlations at both the immediate and delayed post-tests. As seen in Fig. 1, belief in corrected misinformation was correlated highly with memory in both the immediate post-test ($\rho = 0.51$, $p < .001$) and the delayed post-test ($\rho = 0.82$, $p < .001$).[2] While the relationship between memory and belief was significantly stronger at one month than immediately after correction ($Z = -10.61$, $p < .001$, 95% CI = [−0.40, −0.26]), this is likely due to severe floor effects for memory in the immediate post-test restricting the range of beliefs. For facts, memory and belief were similarly highly correlated at the immediate ($\rho = 0.51$, $p < .001$) and delayed post-test ($\rho = 0.85$, $p < .001$), with the delayed post-test correlation being significantly stronger than immediate ($Z = -11.22$, $p < .001$, 95% CI = [−0.41, −0.27]), as seen in Fig. S2.

3.4. The role of memory in belief regression

It is clear that memory for the correction and belief were highly associated at the same time-point. We next wanted to test whether memory for the correction at the delayed time-point explained the degree to which beliefs increased from immediate to one-month post-correction (i.e., belief regression). To this end, we first confirmed that there was a significant belief regression effect by conducting a 2 × 2 between-within ANOVA on misinformation belief with between-subjects factors correction (correction vs. control), and within-subjects factors post-correction retention interval (immediate vs. 1 month). We found an interaction between correction and retention interval, showing that the impact of the correction indeed changed over time ($F(1, 610) = 178.01$; $p < .001$; $MSE = 0.98$; $\eta p^2 = 0.23$). This can be seen in Fig. 2, which shows belief change in the control and correction condition. We found a similar interaction between control vs. affirmation of facts and retention interval, ($F(1, 610) = 195.67$; $p < .001$; $MSE = 0.59$; $\eta p^2 = 0.24$), see Fig. S4.

In order to investigate memory's role in belief regression we created a *belief regression index (b).* We chose this index because it accounts for both the amount that a belief reduces as well as the amount that it rebounds. The equation is as follows, where p = pre-test belief ratings, i = immediate post-test rating, and d = delayed post-test ratings:

$$b = \frac{d - i}{p - i}$$

If belief regressed to the exact rating where it started, the b value would be 1. If belief regression falls between 0 and 1, this indicates that the pre-test remains higher than the delayed post-test; we expected the vast majority of participants to fall between these bounds. If belief regression is above 1, the belief has backfired, with people rating the post-test

[2]For robustness, we also examined this at the item level, collapsed over participants. The immediate post-test was non-significant due to severe floor effects $\rho = 0.09$, $p = .755$; See Supplementary Fig. 3), and the delayed post-test replicated this finding ($\rho = 0.83$, $p < .001$).

above the pre-test. Finally, if participants reduce their belief *even more* after a delay than immediately, the *b* value will be <0.[3]

As can be seen in Fig. 3, collapsed across items, 77.81% of participants exhibited belief regression indices that fell between 0 and 1. Only 11.97% reduced their belief even more in the delayed condition than the immediate condition as shown by participants below 0, and 10.22% increased their belief (or backfired) as shown by those above 1. Next, we examined the extent to which delayed memory accounts for belief regression. We performed a Spearman correlation and found that poorer memory at the delayed post-test was significantly correlated with a greater belief regression index ($\rho = 0.58$, $p < .001$). Note that using a subtraction score capturing the difference between memory at one-month and the immediate post-test produced similar results, $\rho = 0.56$, $p < .001$. We focus on the delayed post-test given that subtraction scores are often less reliable (Cronbach & Furby, 1970).[4]

To better understand the strength of the relationship between delayed post-test memory and the belief regression index, we sought to calculate the theoretical upper bound of the correlation. This is the correlation that would be expected if their true correlation was 1.0, once reliability is taken into account (Schmidt & Hunter, 1996). To determine this, we calculated the reliability of memory at the delayed post-test (Cronbach's alpha = 0.85) as well as the Spearman-Brown-corrected split-half reliability of the belief regression index ($\rho = 0.61$). We found that the upper-bound correlation considering their reliability was 0.72 (geometric mean of 0.61 and 0.85). The relationship between memory and the belief regression index was 0.58/0.72 = 0.81, suggesting that memory at the delayed post-test accounted for 66% of the variance in the belief regression index after correcting for measurement reliability.

For both robustness and converging evidence, we also examined associations between variations in the belief regression index and memory on the item level, collapsed over participants. Using a Spearman correlation, we found that items that were less well remembered had significantly larger belief regression indices ($\rho = 0.65$, $p = .008$). While there were only 16 items and thus findings must be interpreted with caution, this provides additional evidence that memory plays an important role in belief regression. See Fig. S5 for the correlation between belief regression index and memory for affirmed fact items ($\rho = -0.54$, $p < .001$).

We next checked whether age correlated with the belief regression index, given that older adults have previously been shown to have greater belief regression (Swire, Berinsky, et al., 2017). We found that the correlation between age and the belief regression index was non-

[3]We excluded all participants for which this interpretation does not hold true. We removed participants (1) who backfired both immediately *and* after a delay and thus had a belief regression index below zero ($N = 6$), (2) who backfired immediately and then reduced and thus had a belief regression index between 0 and 1 ($N = 2$), and (3) those with belief regression index denominator of 0, since this produces an undefined value ($N = 1$). Note that this involved removing 2.2% of participants in the correction conditions and 1.4% of total data.

[4]We examined whether belief regression was associated with the extent that people had considered the claim. If participants' had spent a good deal of time contemplating it, we would expect them to be more resistant to belief change, both immediately after a correction and in a delayed post-test. We investigated the association between consideration and (1) the belief regression index, (2) belief change from pre-test to immediate post-test, (3) belief change from immediate to delayed post-test, and (4) immediate to delayed post-test memory change. These were all non-signification after the Bonferroni-Holm correction was applied (adjusted $p = .052$, 0.216, 1.00 and 1.00, respectively)

significant (ρ = 0.02, $p = .736$). However, given that we only had four individuals over the age of 65 in our sample, this restriction of age range may have limited our ability to detect a correlation. Finally, we correlated how surprised participants were that the misinformation was false (averaged across items) with the belief regression index. This was to test whether inaccurate information was updated more successfully if an individual is more surprised when it turns out to be false (Butler, Fazio and Marsh, 2011; Metcalfe, 2017). We indeed found that surprise was negatively correlated with the belief regression index (ρ = −0.11, $p = .031$), showing that people who were more surprised immediately after the correction had a lower belief regression index.

3.5. Dissenting vs. forgetting over time

To further investigate whether memory, reduced belief, or both underlies belief regression, we examined belief and memory ratings of each participant for each item. This allowed us to identify the prevalence of different trial types, focusing on dissenting and forgetting. We considered ratings of 0–4 on the 11-point scale as believing/remembering the item to be false, 6–10 as believing/remembering the item to be true, and 5 as unsure. We found that 11.72% of all misinformation items were believed immediately following the correction. This increased to 27.42% after one month, reflecting belief regression over time.

At the immediate and one-month delayed post-tests we identified the prevalence of "dissenting trials," where participants correctly remembered the misinformation as false (i.e., memory ratings 0–4) but still reported believing in the misinformation (i.e., belief ratings 6–10). In other words, participants' memory for the correction is accurate but they are not persuaded and continue to believe in the misinformation. If the number of dissenting trials increase between the immediate and delayed post-tests, this provides evidence for reduced persuasion as a mechanism underlying belief regression. Conversely, if trials increase where participants inaccurately remember that the misinformation was presented as "true" (i.e., memory ratings of 6–10) and report believing in the misinformation (i.e., belief ratings 6–10), this would reflect increased forgetting. If belief regression is due to both failed persuasion and failed memory, we would expect both to increase between the immediate and one-month delay.

Dissenting trials—where misinformation is correctly remembered as false but participants still report believing it—were 7.01% immediately after the correction. This numerically reduced over time, with 6.19% in the delayed post-test ($\chi^2(1, N = 6550) = 3.47$, $p = .062$), although the difference was not significant. If we focus on those who report believing in the misinformation, the dissenters are a sizeable 59.82% at the immediate time point, which reduces to 22.57% at the delayed time point (χ^2 (1, $N = 6550$) = 332.78, $p < .001$). For the full breakdown of the different trial types regarding belief and memory, see Fig. 4 for misinformation and Fig. S6 for facts.

In contrast to the dissenting trials, the proportion of forgetting trials—where participants incorrectly remembered that the misinformation was presented as true and reported believing it—significantly increased from 4.13% in the immediate to 15.59% at the delayed time point ($\chi^2(1, N = 6550) = 483.55$, $p < .001$). When we focus on trials where participants

report believing in the misinformation, the percentage of trials where participants incorrectly recall the misinformation as true increases from 35.24% at the immediate time point, to 56.86% at the delayed time point ($\chi^2(1, N = 6550) = 99.91$, $p < .001$). Together, this suggests that the correction was not perceived to be less persuasive over time, but rather that people's memory for veracity faded. Indeed, 78% of the participants who thought the item was true in the delayed condition had correctly labeled the item as false in the immediate condition.

3.6. Asymmetry in belief regression for misinformation versus facts

Finally, we sought to test whether belief regression is greater for misinformation than facts. We first conducted a 2 × 3 within-subjects ANOVA on belief scores with factors veracity (misinformation vs fact) and pre/post (pre-test, immediate post-test, and delayed post-test). For simplicity we reverse-coded the misinformation items, such that differences in belief regression would reveal themselves as an interaction. We found an interaction of veracity × pre/post, showing that there was indeed asymmetry in belief updating ($F(2, 818) = 76.49$; $p < .001$; $MSE = 0.91$; $\eta p^2 = 0.16$). However, at pre-test the misinformation and facts differed significantly ($t(409) = -13.61$, $p < .001$, $d = -0.85$, 95% CI = [−1.15, −0.86]). This interaction may therefore be due to scaling differences given that the facts were closer to the ceiling than the misinformation was to the floor during initial ratings.

To assess misinformation and facts more equally, we winsorized the items such that the pre-test misinformation was a similar distance to the floor as the pre-test facts were to the ceiling. We thus kept 8 misinformation items ($M = 4.58$) and 8 fact items ($M = 5.49$). This had the added benefit that the misinformation items were assumed to be true the same amount as the fact items were assumed to be false. We reran the same ANOVA on the belief scores constrained to these items, with factors veracity and pre/post. We replicated the previous finding, illustrating a significant veracity × pre/post interaction ($F(2, 818) = 7.22$; $p < .001$; $MSE = 1.19$; $\eta p^2 = 0.02$). However, the effect size was substantially smaller, as can be seen from Fig. 5.

A planned comparison of the winsorized items confirmed that facts were still believed more than the misinformation at the immediate time point ($t(409) = -3.73$, $p < .001$, $d = -0.18$, 95% CI = [−0.46, −0.14]). We next tested whether there was a difference in pre-post increase of misinformation belief and reduction of fact belief. A 2 × 2 within-subjects ANOVA on belief, with factors item veracity (misinformation vs. fact) and retention interval (immediate post-test vs. delayed post-test) revealed a small but significant veracity × retention interval interaction ($F(1, 409) = 4.35$; $p = .038$; $MSE = 0.89$; $\eta p^2 = 0.01$), demonstrating that the misinformation still regressed at a higher rate than the facts for winsorized items. However, the very small effect size must be noted. Indeed, the winsorized belief regression indices of facts (0.37) and misinformation (0.49), respectively ($t(16) = -1.51$, $p = .131$, $d = 0.11$, 95% CI = [−0.28, 0.04]), did not significantly differ.

4. Discussion

The current study found that the association between memory for the correction and belief in the misinformation was very strong: participants with better memory were more likely to

reduce their belief in misinformation, both immediately post-correction ($\rho = 0.51$), and after one month ($\rho = 0.82$). Furthermore, memory at the one-month delayed post-test explained 66% of the variance in belief regression after correcting for measurement reliability. We found that dissenting trials remained stable whereas forgetting trials quadrupled. Indeed, when focusing on cases where misinformation was believed to be true in the delayed post-test, 57% were misremembered as true. Together, this indicates that memory plays a fundamental role in belief regression. While this may seem intuitive, this is the first study to demonstrate this and quantify the strength of the effect. With regards to belief regression asymmetry, we found that corrected misinformation showed more belief regression than affirmed facts, but that this effect was greatly attenuated when items were equated at pre-test.

The importance of memory as a mechanism underlying belief regression aligns well with several previous findings. These include theoretical accounts that attribute the continued influence effect to failures of correct information retrieval (Gilbert, 1991; Sanderson & Ecker, 2020), findings that belief in misinformation is lower when corrections are remembered (Wahlheim et al., 2020), and findings illustrating the relative success of corrective strategies that co-activate the misconception and corrective information in preventing belief regression (Kendeou & O'Brien, 2014). The current results extend these findings by more directly quantifying the relationship between belief regression and memory, and also suggest that individuals who have memory impairments such as from traumatic brain injuries (Burt, Zembar, & Niederehe, 1995), psychiatric disorders (Kinsella et al., 1996), or age-related memory decline (e.g., decreased recollection; see Brashier & Schacter, 2020, for a review) may be the most susceptible to believing misinformation.

Post-correction dissenting was still an important component of why people did not change their minds at the immediate and one-month time point. Although dissenting trials were only 7% immediately and 6% after one month, these would likely increase if the corrections were not citing reputable sources or providing sufficient supporting evidence. Notably, dissenting trials did not increase between the immediate and delayed post-test. The current study suggests that belief is unlikely to increase independently of the recollection of the false tag. This aligns with research showing that people with vested interests in the misinformation being true are not more susceptible to belief regression (Swire, Berinsky, et al., 2017). In other words, there is limited evidence for "motivated forgetting," where people increase belief faster over time if the correction counters their worldview.

Regarding belief regression asymmetry, we replicated the finding that corrected misinformation shows more belief regression than affirmed facts (Skurnik et al. 2007; Swire, Berinsky, et al., 2017). However, this effect was greatly attenuated when items were equated at pre-test, suggesting that much of this asymmetrical effect found in previous reports is due to scaling issues (e.g., regression to the mean). Although we found that misinformation still regressed at a higher rate than facts after items were winsorized, we must question how meaningful and generalizable this finding is to the real-world given the small effect size ($\eta p^2 = 0.01$). A prime take-away is that future research wishing to investigate asymmetrical updating of misinformation and facts should ensure that items are equivalent at pre-test.

This research opens several new areas for inquiry regarding the nature of the memory mechanisms involved in belief regression. For instance, future research could more explicitly investigate deficits of binding misinformation to the false tag (Zimmer, Mecklinger, & Lindenberger, 2006) or binding other corrective components such as the reputable source or associative details for why the information is false. Research could also explicitly test the principles of Kendeou and O'Brien's (2014) knowledge revision framework (encoding, passive activation, co-activation, integration, and competing activation) to further understand which process(es) best accounts for belief regression. There might also be other mechanisms for belief regression beyond memory, such as changes in demand characteristics (where participants attempt to predict and comply with researchers' expectations) and expressive responding (where participants report believing in misinformation to communicate something else to the researcher, such as their political viewpoint; Schaffner & Luks, 2018). Furthermore, future studies should separately measure belief in the correction as well as belief in the misinformation to gain a more complete understanding of misinformation processing. Finally, future research could be extended by measuring memory and belief at three time points rather than only two. This would allow for forgetting curves to be fitted and would provide greater predictive power.

Though the current results are compelling, a limitation of this study is that it was conducted in the general population, and thus cannot be generalized to conspiracy theorists or other extreme populations of interest. Population-specific studies with beliefs specific to those individuals (for instance, COVID-19 vaccine beliefs in an anti-vaccine cohort) would be useful. Furthermore, this study should be replicated with stimuli that are more emotive, self-relevant, or integral to peoples' worldviews. Although we found that people who had considered misinformation more deeply were not more or less prone to belief change or belief regression, this was with using relatively non-emotive stimuli. Finally, given that the current participants were asked to rate their surprise in each item to ensure sufficient encoding of corrections and affirmations, we may have underestimated forgetting given that real-world corrections are likely to be processed more shallowly.

In sum, it is not only important to understand what drives initial belief change, but also how it can be sustained over time. This study finds that memory failure plays a substantial role in beliefs regressing back to pre-correction levels over a one-month period. Regarding real-world implications, fact checkers should aim to improve the memorability of their corrections in order to prevent memory failures. This can be done with known memory enhancers such as the repetition of corrections (Toppino, Kasserman, and Mracek, 1991) or encouraging people to read carefully to increase depth of encoding (Moscovitch & Craik, 1976). While the vast majority of current research is conducted with no retention interval (Dias & Sippitt, 2020), or does not measure memory at all, the current findings clearly underscore the importance of measuring belief regression over time and taking into account the powerful effects of forgetting. Considering these factors in future studies will enable a deeper understanding of why people believe in misinformation as well as provide effective novel intervention approaches.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

This study was funded by a National Institute of Health Pathway to Independence Award (1K99CA248720-01A) to BST and a grant of the Volkswagen Foundation (Initiative "Artificial Intelligence and the Society of the Future") to BST.

Data availability

We have made our data available at https://doi.org/10.5061/dryad.2rbnzs7r9.

References

Berinsky AJ (2017). Rumors and health care reform: Experiments in political misinformation. British Journal of Political Science, 47(2), 241–262. 10.1017/S0007123415000186

Brashier NM, & Schacter DL (2020). Aging in an era of fake news. Current Directions in Psychological Science, 29(3), 316–323. [PubMed: 32968336]

Burt DB, Zembar MJ, & Niederehe G. (1995). Depression and memory impairment: A meta-analysis of the association, its pattern, and specificity. Psychological Bulletin, 117(2), 285–305. 10.1037/0033-2909.117.2.285 [PubMed: 7724692]

Butler AC, Fazio LK, & Marsh EJ (2011). The hypercorrection effect persists over a week, but high-confidence errors return. Psychonomic Bulletin & Review, 18(6), 1238–1244. [PubMed: 21989771]

Carey JM, Guess AM, Loewen PJ, Merkley E, Nyhan B, Phillips JB, & Reifler J. (2022). The ephemeral effects of fact-checks on COVID-19 misperceptions in the United States, Great Britain and Canada. Nature Human Behaviour, 1–8.

Cronbach LJ, & Furby L. (1970). How we should measure" change": Or should we? Psychological Bulletin, 74(1), 68.

Dennis N, Gutchess A, & Thomas A. (2020). Overview of models of cognitive aging. In Thomas A, & Gutchess A. (Eds.), The Cambridge handbook of cognitive aging: A life course perspective (Cambridge handbooks in psychology) (pp. 5–31). Cambridge: Cambridge University Press. 10.1017/9781108552684.002.

Dias N, & Sippitt A. (2020). Researching fact checking: Present limitations and future opportunities. The Political Quarterly, 91(3), 605–613. 10.1111/1467-923X.12892

Faul F, Erdfelder E, Buchner A, & Lang A-G (2009). Statistical power analyses using G*power 3.1: Tests for correlation and regression analyses. Behavior Research Methods, 41(4), 1149–1160. 10.3758/BRM.41.4.1149 [PubMed: 19897823]

Gallo DA, Bell DM, Beier JS, & Schacter DL (2006). Two types of recollection-based monitoring in younger and older adults: Recall-to-reject and the distinctiveness heuristic. Memory, 14(6), 730–741. [PubMed: 16829489]

Gilbert DT (1991). How mental systems believe. American Psychologist, 46(2), 107.

Hill SJ, Lo J, Vavreck L, & Zaller J. (2013). How quickly we forget: The duration of persuasion effects from mass communication. Political Communication, 30(4), 521–547.

Hovland CI, Lumsdaine AA, & Sheffield FD (1949). Experiments on mass communication. (Studies in social psychology in World War II). 3.

Kendeou P, & O'Brien EJ (2014). The knowledge revision components (KReC) framework: Processes and mechanisms. In Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences (pp. 353–377). Boston Review. 10.7551/mitpress/9737.001.0001.

Kinsella G, Murtagh D, Landry A, Homfray K, Hammond M, O'Beirne L, … Ponsford J. (1996). Everyday memory following traumatic brain injury. Brain Injury, 10(7), 499–508. 10.1080/026990596124214 [PubMed: 8806010]

Kowalski P, & Taylor AK (2017). Reducing students' misconceptions with refutational teaching: For long-term retention, comprehension matters. Scholarship of Teaching and Learning in Psychology, 3(2), 90–100. 10.1037/stl0000082

Metcalfe J. (2017). Learning from errors. Annual Review of Psychology, 68, 465–489.

Moscovitch M, & Craik FIM (1976). Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. Journal of Verbal Learning and Verbal Behavior, 15(4), 447–458. 10.1016/S0022-5371(76)90040-2

O'Rear AE, & Radvansky GA (2020). Failure to accept retractions: A contribution to the continued influence effect. Memory & Cognition, 48(1), 127–144. 10.3758/s13421-019-00967-9 [PubMed: 31317393]

Peer E, Brandimarte L, Samat S, & Acquisti A. (2017). Beyond the turk: Alternative platforms for crowdsourcing behavioral research. Journal of Experimental Social Psychology, 70, 153–163. 10.1016/j.jesp.2017.01.006

Peter C, & Koch T. (2016). When debunking scientific myths fails (and when it does not) the backfire effect in the context of journalistic coverage and immediate judgments as prevention strategy. Science Communication, 38(1), 3–25. 10.1177/1075547015613523

Rich PR, Van Loon MH, Dunlosky J, & Zaragoza MS (2017). Belief in corrective feedback for common misconceptions: Implications for knowledge revision. Journal of Experimental Psychology: Learning, Memory, and Cognition, 43(3), 492–501. 10.1037/xlm0000322 [PubMed: 27762579]

Rich PR, & Zaragoza MS (2020). Correcting misinformation in news stories: An investigation of correction timing and correction durability. Journal of Applied Research in Memory and Cognition, 9(3), 310–322. 10.1016/j.jarmac.2020.04.001

Sanderson JA, & Ecker UKH (2020). The challenge of misinformation and ways to reduce its impact. In Handbook of learning from multiple representations and perspectives. Routledge.

Schacter DL (2008). Searching for memory: The brain, the mind, and the past. Basic Books.

Schaffner BF, & Luks S. (2018). Misinformation or expressive responding? What an inauguration crowd can tell us about the source of political misinformation in surveys. Public Opinion Quarterly, 82(1), 135–147.

Schmidt FL, & Hunter JE (1996). Measurement error in psychological research: Lessons from 26 research scenarios. Psychological Methods, 1(2), 199–223.

Schwarz N, Sanna LJ, Skurnik I, & Yoon C. (2007). Metacognitive experiences and the intricacies of setting people straight: Implications for Debiasing and public information campaigns. In, Vol. 39. Advances in experimental social psychology (pp. 127–161). Elsevier. 10.1016/S0065-2601(06)39003-X.

Seifert CM (2002). The continued influence of misinformation in memory: What makes a correction effective?. In psychology of learning and motivation (Vol. 41, pp. 265–292). Academic Press.

Swire B, Berinsky AJ, Lewandowsky S, & Ecker UKH (2017). Processing political misinformation: Comprehending the trump phenomenon. Royal Society Open Science, 4(3). 10.1098/rsos.160802

Swire B, Ecker UKH, & Lewandowsky S. (2017). The role of familiarity in correcting inaccurate information. Journal of Experimental Psychology. Learning, Memory, and Cognition. 10.1037/xlm0000422

Swire-Thompson B, Miklaucic N, Wihbey JP, Lazer D, & DeGutis J. (2022). The backfire effect after correcting misinformation is strongly associated with reliability. Journal of Experimental Psychology. General, 9(3), 286–299.

Toppino TC, Kasserman JE, & Mracek WA (1991). The effect of spacing repetitions on the recognition memory of young children and adults. Journal of Experimental Child Psychology, 51(1), 123–138. 10.1016/0022-0965(91)90079-8 [PubMed: 2010724]

Wagenmakers EJ, Marsman M, Jamil T, et al. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. Psychonomic Bulletin & Review, 25, 35–57. 10.3758/s13423-017-1343-3 [PubMed: 28779455]

Wahlheim CN, Alexander TR, & Peske CD (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. 10.1177/0956797620952797 [PubMed: 32976064]

Yonelinas AP (2002). The nature of recollection and familiarity: A review of 30 years of research. Journal of Memory and Language, 46(3), 441–517. 10.1006/jmla.2002.2864

Zacks RT, Radvansky G, & Hasher L. (1996). Studies of directed forgetting in older adults. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22(1), 143. [PubMed: 8648283]

Zimmer H, Mecklinger A, & Lindenberger U. (2006). Handbook of binding and memory: Perspectives from cognitive neuroscience. Oxford University Press.

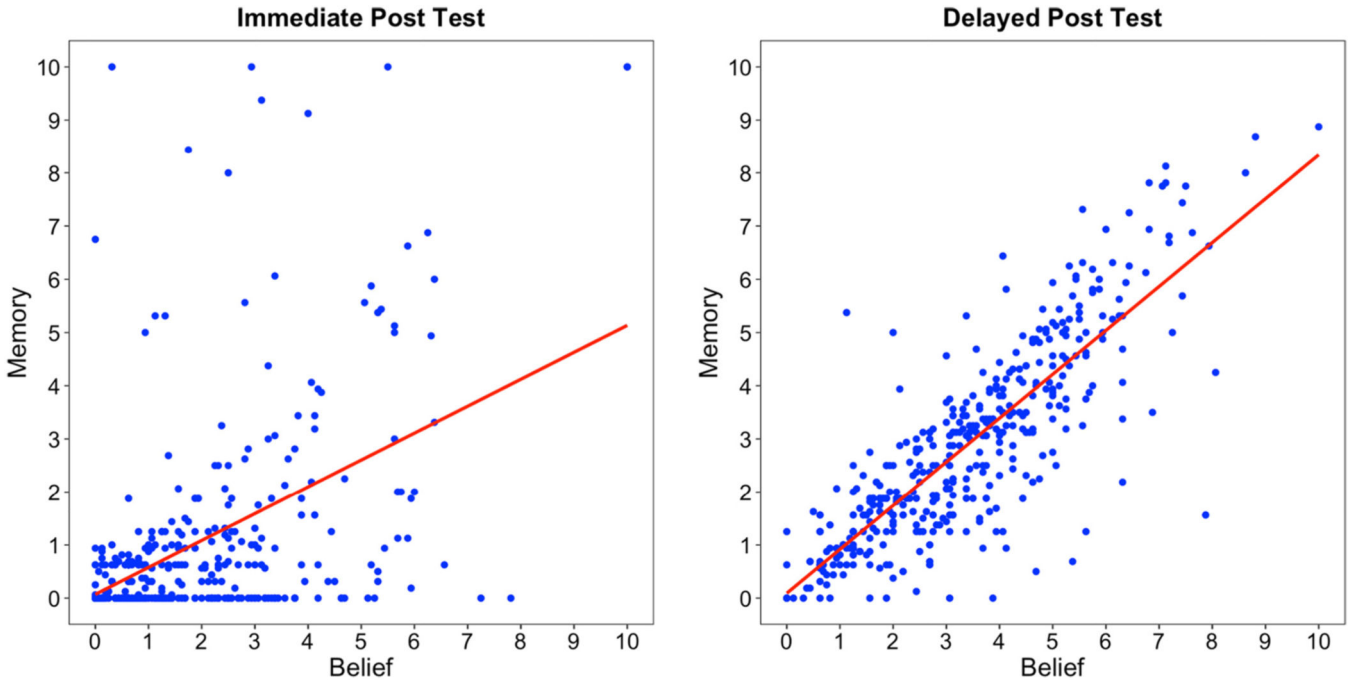

Fig. 1.

Correlation between participants' memory for correction veracity and belief in misinformation (collapsed across items) immediately after the correction (left), and one month after the correction (right).

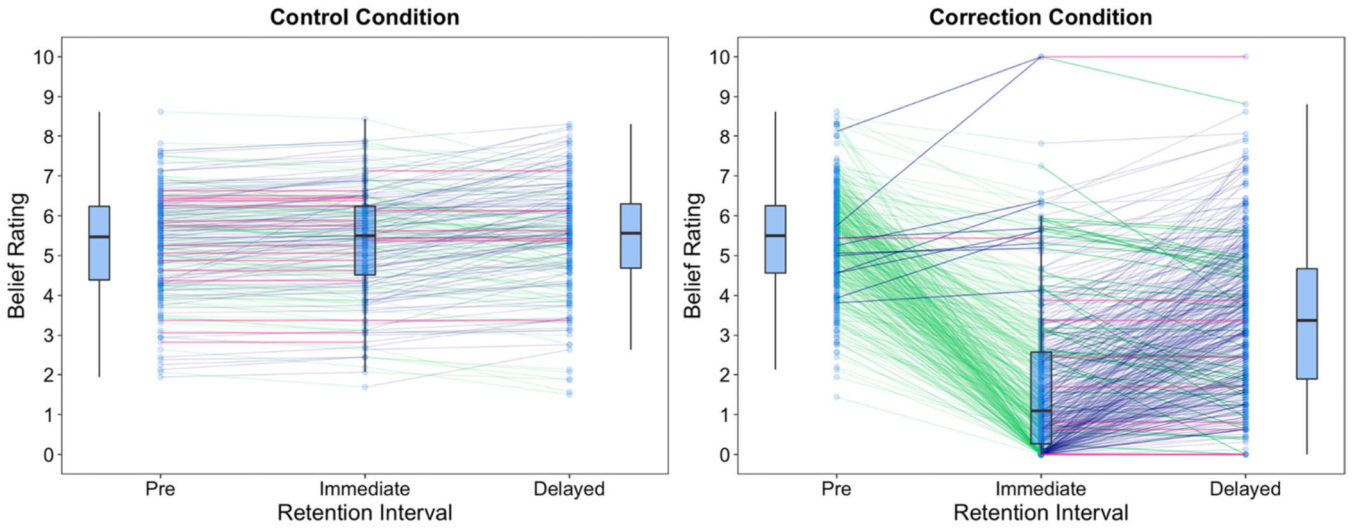

Fig. 2.

Average misinformation belief per participant at pre-test, immediate post-test, and delayed post-test. In both control (left) and correction (right) conditions, green indicates a reduction of belief, dark blue indicates increasing belief, and pink lines indicate no change.

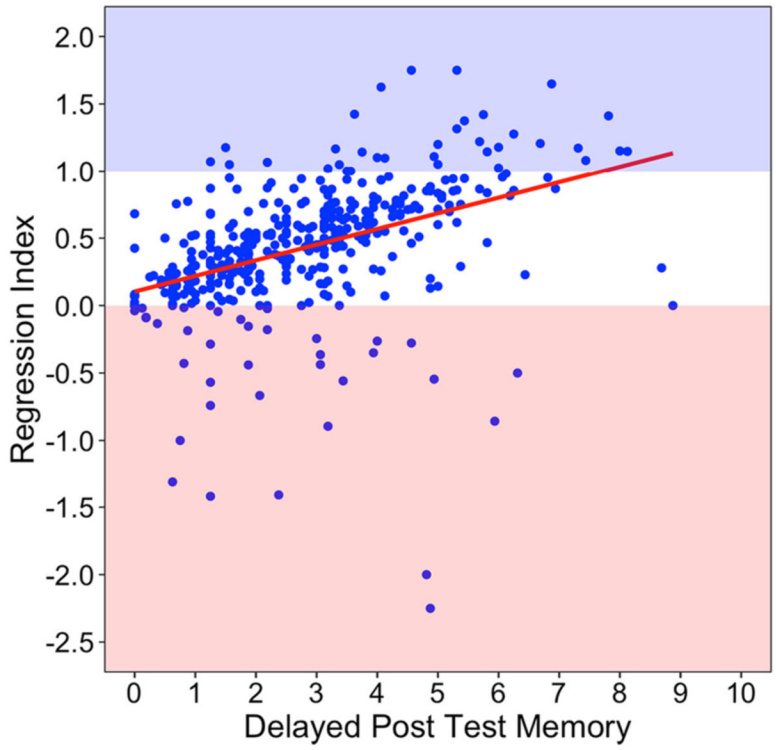

Fig. 3.

Correlation between belief regression index and memory for whether the misinformation is false in the delayed post-test ($\rho = 0.58$, $p < .001$). Participants who backfired (shaded blue area), demonstrated belief regression (the white area), or reduced their belief even more after a delay than immediately after corrections (shaded pink area).

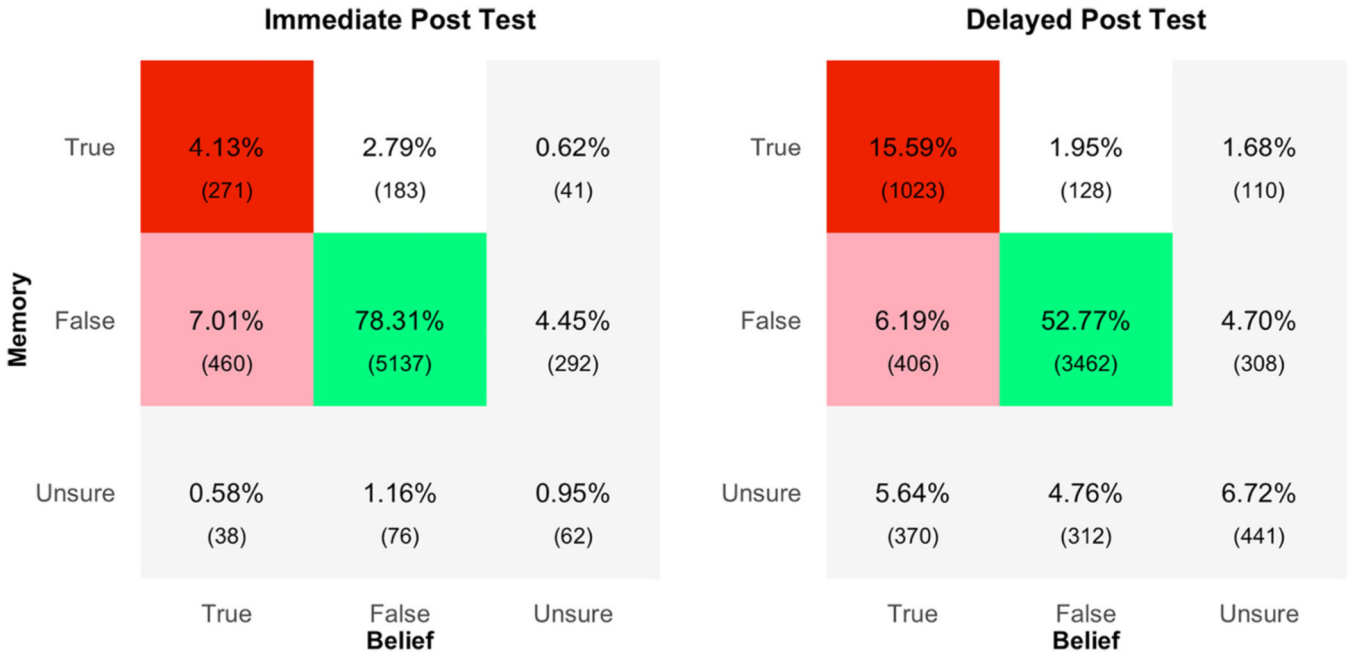

Fig. 4.

Belief in misinformation and memory for correction veracity broken down by percent trial type across all items and participants for immediate post-test (left) and delayed post-test (right). Raw trial counts in parentheses. Rating of 0–4 = false, 5 = unsure, 6–10 = true. The green cells indicate correct trials (accurate memory and disbelief in misinformation), the pink cells indicate dissenting trials (accurate memory but belief in misinformation), and the red cells indicate forgetting trials (inaccurate memory and belief in misinformation).

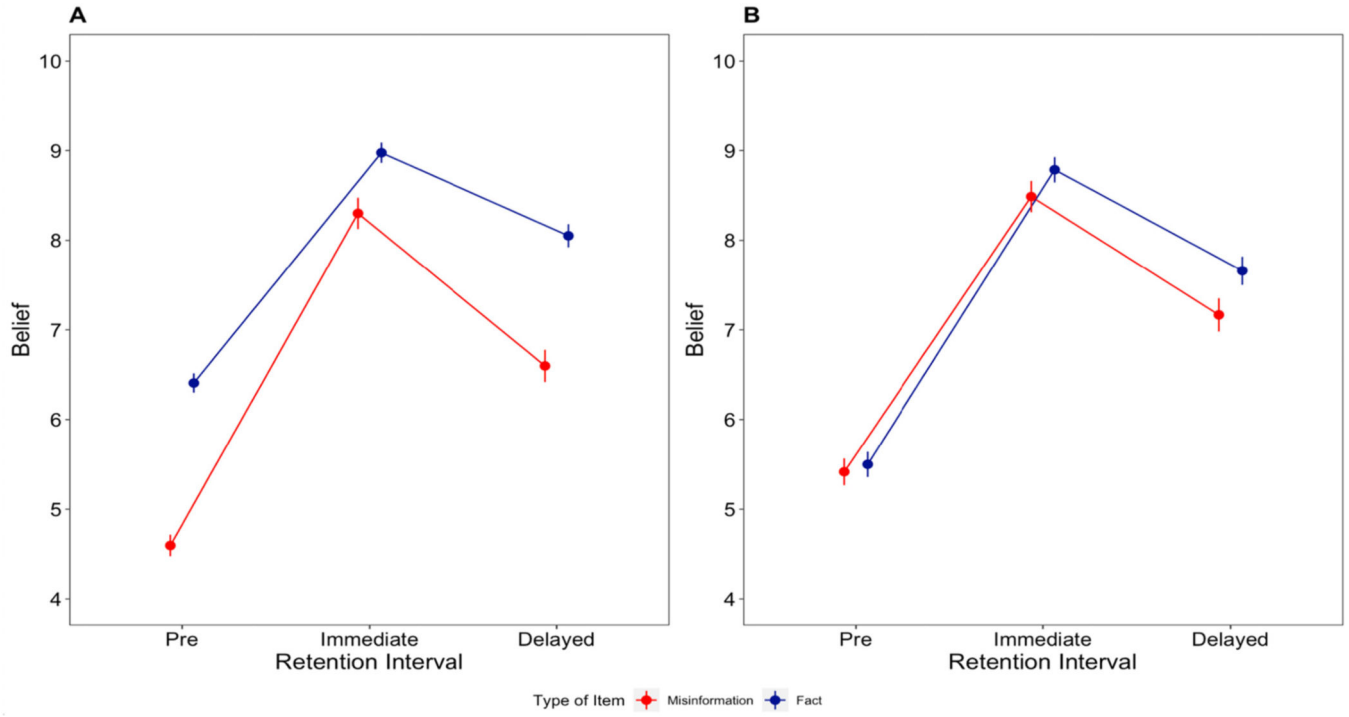

Fig. 5.

Belief in misinformation (red; reverse-coded) and facts (blue) pre and post corrections/affirmations. Panel A (left) shows all items and panel B (right) shows the winsorized items.

Table 1

Example misinformation and correction.

Misinformation	Correction
Mercury in vaccines can cause harm	Mercury in vaccines can cause harm **This is false** There are two types of mercury. Methyl mercury builds up in the body and is toxic. Ethyl mercury—the type within vaccines—is excreted rapidly from the body. In 2006, an expert panel assembled by the World Health Organization concluded that there was “no evidence of toxicity in infants, children or adults exposed to [mercury] in vaccines”.