The Role of Familiarity in Correcting Inaccurate Information

Briony Swire[*1], Ullrich K. H. Ecker[2], and Stephan Lewandowsky[3]

in press: Journal of Experimental Psychology: Learning, Memory, and Cognition.

[*] Corresponding author

[1] School of Psychology, University of Western Australia (M304), Perth 6009, Australia, briony.swire-thompson@research.uwa.edu.au

[2] School of Psychology, University of Western Australia (M304), Perth 6009, Australia, ullrich.ecker@uwa.edu.au

[3] School of Experimental Psychology, University of Bristol (12a Priory Road, Bristol BS8 1TU, UK), and School of Psychology, University of Western Australia (M304), Perth 6009, Australia. stephan.lewandowsky@bristol.ac.uk

Acknowledgments: This research was facilitated by a Fulbright Postgraduate Scholarship from the Australian-American Fulbright Commission and a University Postgraduate Award from the University of Western Australia to the first author, and a Discovery Grant from the Australian Research Council to the second and third author. The third author was supported by a Wolfson Research Merit Fellowship from the Royal Society while this work was conducted. We thank Charles Hanich for research assistance. The lab web address is http://www.cogsciwa.com.

Abstract

People frequently continue to use inaccurate information in their reasoning even after a credible retraction has been presented. This phenomenon is often referred to as the continued influence effect of misinformation. The repetition of the original misconception within a retraction could contribute to this phenomenon, as it could inadvertently make the "myth" more familiar—and familiar information is more likely to be accepted as true. From a dual-process perspective, familiarity-based acceptance of myths is most likely to occur in the absence of strategic memory processes. We thus examined factors known to affect whether strategic memory processes can be utilized. Participants rated their belief in various statements of unclear veracity, and facts were subsequently affirmed and myths were retracted. Participants then re-rated their belief either immediately or after a delay. We compared groups of young and older participants, and we manipulated the amount of detail presented in the affirmative/corrective explanations, as well as the retention interval between encoding and a retrieval attempt. We found that (1) a greater level of explanatory detail promoted more sustained belief change, and (2) fact affirmations promoted more sustained belief change in comparison to myth retractions over the course of one week (but not over three weeks), particularly for older adults. This supports the notion that familiarity is indeed a driver of continued influence effects.

Keywords: misinformation; continued influence effect; belief updating; familiarity backfire effect; older adults; dual-process models of memory.

The Role of Familiarity in Correcting Inaccurate Information

Every day we process an extraordinary amount of information, and it is often up to the individual to discern fact from fiction. A proportion of this information is inevitably inaccurate and deserves to be corrected after initial encoding. In order to maintain an accurate and up-to-date representation of the world, ideally we would disregard invalidated information. However, we are far from perfect at performing this task, as corrected misinformation often continues to influence memory and reasoning—this persistence is termed the *continued influence effect* of misinformation (cf. Ecker, Lewandowsky, Swire, & Chang, 2011; Johnson & Seifert, 1994). With "fake news" fast becoming a global issue, and with the increased spread of misinformation over social media, the ability to effectively correct misinformation has never been more important (Connolly et al., 2016; Lavoipierre, 2017). From relatively benign misconceptions such as "ostriches hide their heads in the sand" to more malignant misinformation such as "the MMR vaccine causes autism" (Poland & Spier, 2010), studies have consistently observed a continued influence effect. In other words, simply stating that information is incorrect is often ineffective, with simple retractions typically only halving the number of references to a critical piece of misinformation relative to a no-retraction control condition (see Lewandowsky, Ecker, Seifert, Schwarz, & J. Cook, 2012, for a review). Part of the reason why corrections are often ineffective may arise because corrections typically repeat the misinformation, thereby making it more familiar. The present paper investigates whether the continued influence effect is at least partially familiarity-driven, and how beliefs change over time after a credible correction has been presented.

The Illusory Truth Effect

The *illusory truth effect* occurs when increased familiarity gives rise to the illusion that information is valid and inadvertently increases an individual's belief (Begg, Anas, & Farinacci, 1992; Dechene, Stahl, Hansen, & Wanke, 2010; Wang, Brashier, Wing, Marsh, & Cabeza, 2016). For example, DiFonzo, Beckstead, Stupak, and Walders (2016) presented participants with rumors about campus life, such as a professor giving a student good grades to keep them quiet about the professor's plagiarism. DiFonzo and colleagues varied the number of presentations (from 0-9 times) and found that belief in the statements was logarithmically related to the number of repetitions, such that belief increased with each repetition (albeit in diminishing amounts). In line with this, a recent popular news survey found that 75 % of people assumed fake headlines to be true if they were familiar (Silverman, 2016).

The illusory truth effect could be problematic when attempting to correct misinformation, as a correction often repeats the original claim. For example, truthfully stating that playing Mozart to your child will *not* boost its IQ mentions the two concepts of "Mozart" and "increased IQ", thereby making the link between the concepts more familiar even though the statement seeks to dispel the Mozart-IQ myth. This inadvertent increase in familiarity may reduce the effectiveness of a correction and may thus contribute to the continued influence effect of misinformation.

Strategic and Automatic Memory Processes

The potential familiarity-related difficulties that arise during the correction of misinformation may be explained from a dual-processing perspective. Dual-process theories of memory assume a dichotomy between automatic memory processes, which include familiarity, and strategic memory processes such as recollection and output monitoring (cf. Brown &

Warburton, 2006; Diana, Yonelinas, & Ranganath, 2007; Rugg & Curran, 2007; Yonelinas, 2002; Yonelinas & Jacoby, 2012; Zimmer & Ecker, 2010). Familiarity is thought to be a fast, context-free automatic process that allows for the rapid recognition of previously encountered information. Recollection, by contrast, is a slower process thought to allow for the retrieval of contextual details, such as the information's source, its spatiotemporal encoding context, or its veracity. In the case of corrected misinformation, it is often assumed that a "negation tag" is linked to the original statement, for example, "playing Mozart to your child will boost its IQ—*NOT TRUE*" (Gilbert, Krull, & Malone, 1990). Thus, a corrected statement may require strategic memory processes for veracity to be successfully retrieved, as the negation tag is at risk of being lost if only automatic processes are employed, which may however identify the statement (e.g., the Mozart-IQ link) as familiar.

Regardless of whether statements are correct or have been invalidated, existing memory representations will be activated in response to cues via automatic retrieval to the extent that the information is familiar (cf. Ayers & Reder, 1998). To avoid reliance on familiar but invalid information, strategic memory processes are required to act as a filter of automatically retrieved memory output. However, strategic memory processes take effort and often fail (e.g., Herron & Rugg, 2003), and thus people can rely upon invalid but automatically retrieved information in their judgments (Ecker et al., 2011; Koutstaal & Schacter, 1997; Reyna & Lloyd, 1997; Roediger, Watson, McDermott, & Gallo, 2001). A post-correction misinformation effect is therefore likely to occur when misinformation has been automatically activated but strategic memory processes have failed (Ecker, Lewandowsky, & Tang, 2010). Familiarity can thus hinder the remediating effect of a correction when the repetition of misinformation in the course

of its correction boosts an invalid item's familiarity such that it outweighs the correction's strategic-retrieval dependent corrective effect.

The Familiarity Backfire Effect

Some reports suggest that the familiarity boost associated with a correction can be so detrimental that it causes a *familiarity backfire effect*, such that a correction can ironically increase an individual's belief in the very misconception the correction is aiming to rectify (J. Cook & Lewandowsky, 2011; Lewandowsky et al., 2012).[1] An unpublished manuscript by Skurnik, Yoon, and Schwarz (2007; as cited in Schwarz, Sanna, Skurnik, & Yoon, 2007), is frequently cited when discussing the familiarity backfire effect (e.g., Berinsky, 2015; J. Cook, Bedford, & Mandia, 2014; Gemberling & Cramer, 2014; Lilienfeld, Marshall, Todd, & Shane, 2015; Peter & Koch, 2016; Schultz, 2012). Skurnik et al. presented participants with a flyer presenting a number of flu-vaccine related claims. The flyer was either presented in a "myths vs. facts" format, which affirmed the factual statements and refuted the incorrect, or the flyer only affirmed the facts but did not mention the myths (or, in a control condition, there was no flyer at all). Immediately after participants read the "myth vs. facts" flyer, they were able to reliably distinguish between myths and facts, yet after a delay of 30 minutes, participants misidentified 15% of the myths as facts (compared to only 2 % of the facts being classified as false). Also, after a 30-minute delay, participants in the myths vs. facts condition had less favorable attitudes towards the vaccine than participants who had never seen the flyer at all.

[1] The term familiarity backfire effect has been used somewhat inconsistently. The term is sometimes used simply when myths are misremembered as facts, without a control condition or baseline comparison (cf. Peter & Koch, 2016). However, we argue it should only pertain to cases where a correction inadvertently *increases* myth belief relative to a no-correction or pre-correction baseline. Our definition is in accordance with other backfire effects such as the *worldview backfire effect* (Lewandowsky et al., 2012).

Skurnik et al.'s (2007) finding that people had a less favorable attitude towards vaccines than those who did not view the flyer may reflect a familiarity backfire effect. However, given that Skurnik et al. only focused upon one contentious issue, there is an alternative explanation—namely a *worldview backfire effect*. This backfire effect ensues when a correction challenges a person's belief system and the person becomes motivated to defend their worldview, resulting in an increased belief in the inaccurate information, relative to a situation where the correction was never presented (Lewandowsky et al., 2012; Trevors, Muis, Pekrun, Sinatra, & Winne, 2016). This effect is a known risk when debating contentious issues and can contribute to belief polarization (Hart & Nisbet, 2012; Nyhan & Reifler, 2010). Nyhan and colleagues (Nyhan, Reifler, Richey, & Freed, 2014; Nyhan & Reifler, 2015) recently demonstrated that corrections of vaccine-related misconceptions can backfire in people concerned about vaccination safety and/or opposed to vaccinations. Thus, when attempting to measure the effects of familiarity, it would be beneficial to not exclusively focus upon politicized information (e.g., Berinsky, 2015) or contentious topics such as vaccination to avoid confounding the effects of familiarity and worldview.

Regarding the misidentification of myths as facts in the Skurnik et al. (2007) study, results from the comparison between the "myths vs. facts" condition and the no-flyer control condition are not available, so it is unclear whether myth belief was *greater* after presenting corrections than after not presenting any information at all. However, misremembering myths as facts could in general reflect an interesting hurdle in belief updating, even if there is no "true" backfire effect. Considering the benefits of clear communication, in particular in the context of myth debunking, it is important to better understand the factors that may contribute to this

differential forgetting of myths relative to facts (or more precisely, differential forgetting regarding the veracity of myths relative to facts).

Theoretically, if people are presented with explanations affirming facts or refuting myths, belief in facts may be sustained over time, whereas myth items appear to be "forgotten", simply because automatic and strategic memory processes operate in concert for facts but stand in opposition for myths (cf. Brainerd & Reyna, 2008; Jacoby, 1991; Toth, 1996). For fact items, regardless of whether automatic processes or strategic memory processes are employed, both would lead a participant to conclude that the item is true. By contrast, if a participant is unable to correctly recall the correction of a myth due to the forgetting that primarily affects strategic memory processes, the familiarity of the myth—boosted by its repetition during the correction—could lead to the myth being inaccurately accepted as true.

Factors Likely to Influence the Correction of Information: Detail, Time, and Age

It follows from the dual-process notion that the relative impact of familiarity on corrections could potentially be influenced by factors that are known to affect strategic memory processes, including (1) the amount of detail presented in the corrective explanation, (2) the retention interval between encoding and a retrieval attempt, and (3) the age of the participant.

Regarding the correction's level of detail, providing sufficiently detailed explanations as to why a piece of misinformation is false—in other words, providing a detailed *refutation* rather than a sparse "that-is-not-true" *retraction* (cf. Guzzetti, 2000)—might counteract familiarity's influence. For example, where a simple retraction would merely state that "listening to Mozart will *not* boost your child's IQ", a detailed refutation would also explain why (e.g., by highlighting that scientific findings were misrepresented in a popular yet unscientific book, and that the original study neither tested infants nor measured IQ; Campbell, 1997; Pietschnig,

Voracek, & Formann, 2010; Rauscher, Shaw, & Ky, 1993). Thus, refutations directly address the misconception and explain the reasons why the misinformation is false and/or where the misconception originated. Refutations have been found to promote belief change over long periods of time (Diakidoy, Mouskounti, Fella, & Ioannides, 2016; Guzzetti, Snyder, Glass, & Gamas, 1993); it is assumed that refutations are more effective than retractions because they encourage the detection of inconsistencies between a person's inaccurate beliefs and the corrective information, and because they provide richer informational detail that can later support recollection of the correction (Guzetti, 2000).

Regarding the retention interval, failure of strategic processes is particularly likely when there is some delay between encoding and attempted retrieval, as strategic recollection of details diminishes with time, whereas familiarity stays relatively constant (Knowlton & Squire, 1995). Thus, false acceptance of myths based on their familiarity seems particular likely at longer retention intervals.

Regarding age, older adults have less efficient strategic memory processes than young adults, whereas automatic processing such as familiarity-detection remains relatively age-invariant (e.g., Prull, Dawes, Martin, Rosenberg, & Light, 2006). In particular, older adults seem to become less efficient at binding item and context information (Naveh-Benjamin, 2000); therefore, the mnemonic link between a statement and its veracity could be weaker in older adults. This is in line with the finding that source memory—memory for where or how information was acquired—is particularly susceptible to the effects of ageing (e.g., Glisky, Rubin, & Davidson, 2001). Consistent with this notion, Skurnik, Yoon, Park, and Schwarz (2005) found that older adults were particularly likely to misremember myths as facts after repeated retractions (compared to single retractions) after a three-day retention interval (but not

after 30-minutes, and not in younger adults as per the Skurnik et al. 2007 study). However, it is difficult to draw firm conclusions from the Skurnik et al. (2005) study for several reasons: (i) there was no control group where corrections were not presented at all or pre-manipulation belief ratings were measured; (ii) no cognitive screening task was given to participants, potentially reducing the generalizability of findings; and (iii) health claims were used that were arbitrarily labeled as valid or invalid without explanation, even though all claims were actually true—thus some corrections were misleading, and distrust in the corrections may have contributed to the results, as it is well established that source credibility is an influential factor in the persuasiveness of messages (Eagly & Chaiken, 1993; Guillory & Geraci, 2013).

In summary, factors such as the correction's level of explanatory detail, retention interval, and participant age are likely to play a role in determining the success of a correction, but their specific importance is unclear and findings have been inconsistent. By manipulating and comparing these factors, the present research aimed to clarify if and under what conditions familiarity is most problematic. Experiment 1 tested young adults, Experiment 2 tested older adults.

Experiment 1

This study presented an undergraduate population with both incorrect and correct claims (i.e., myths and facts), then corrected the false claims in a way that boosted their familiarity. To this end, participants were presented with a range of statements of unclear veracity that were subsequently labeled as true or false. People's belief in the statements was measured both before the true/false explanation and in a post-manipulation test phase to yield a measure of belief change. To avoid the problems associated with posttest-only designs (Morris, 2008), we used a

pretest-posttest design so that each individual could be used as their own control (Hunter & Schmidt, 2004).

Level of explanatory detail and study-test retention interval were manipulated to identify the parameters of corrections that promote successful discounting of misinformation. The experiment used a 2 × 2 × 3 within-between design, with within-subjects factors type of item (myth vs. fact) and type of explanation (the veracity of each statement was explained either briefly or in some detail), and the between-subjects factor retention interval (immediate, 30-minute, or one-week). In some studies, continued influence effects were found primarily in more indirect measures of belief that require participants to use the misinformation in reasoning (cf. Johnson & Seifert, 1994). Therefore, inference questions were also administered at test, serving as a more indirect measure of belief that could help avoid issues related to social desirability.

We hypothesized that (1) detailed explanations would lead to stronger belief change than brief explanations for both myths and facts, and (2) belief change would be more sustained over time after fact affirmation compared to myth retraction, as false familiarity-based acceptance of myths would seem particular likely at longer retention intervals. We did not expect a backfire effect, as there are no clear demonstrations of a true familiarity backfire effect in the peer-reviewed literature. However, we hypothesized that one was theoretically most likely to occur with a brief retraction after a one-week delay.

Method

Participants. A power analysis (conducted with GPower3; Faul, Erdfelder, Lang, & Buchner, 2007) suggested that 78 participants were required in order to detect a small-to-medium effect (effect size $f = .15$) with $\alpha = .05$, $1 - \beta = .80$, and a moderate correlation between repeated measures of $r = .50$. Participants were 100 undergraduate students from the University of

Western Australia who volunteered after reading an ethically-approved information sheet. They received course credit for participation. Two participants did not complete the study, and five participants over the age of 30 were excluded as age outliers. This conformed to the age split of prior research (Skurnik et al., 2005) as well as the outlier labeling rule threshold (i.e., 2.2 times the interquartile range above the third quartile [Hoaglin & Iglewicz, 1987]), which was 29.8 years of age. The final pool included $N = 93$ participants, with 19 males and 74 females between 16-28 years of age, and a mean age of $M = 19.11$ ($SD = 2.53$).

Stimuli. There were 20 myths and 20 facts, each with a corresponding brief explanation, a detailed explanation, and two inference questions. An example myth / fact and the corresponding explanations and example inference questions are given in Table 1 (see Supplement A for the complete list of items, explanations, and inference questions). Brief explanations simply stated whether the item was a myth or a fact with no further clarification. They explicitly repeated the initial statement twice (once in the original and once in a negated format if the item was a myth). Thus participants encountered the initial statement three times altogether: once when being initially rated, and twice in the explanation. Detailed explanations also provided the myth/fact label but in addition included three or four sentences of further information; myth retractions did not provide a causal alternative to the myth but rather explained why the myth was wrong and/or where it originated from. Detailed explanations explicitly repeated the initial statement only once, but elements of the statement were repeated in the additional information.

Inference questions were rated on an 11-point scale, with the specific scale-value range dependent on the item; for example, some items were rated on a 0-10 scale, others were rated on a 0-20 % scale with 2 % increments.

Two pilot studies were conducted to select stimuli from a list of 80 items (55 myths and 25 facts) that was initially compiled by selecting various items from websites such as New Scientist, Scientific American, and myth busting programs such as QI. Each item was researched to the best of our ability, and where possible evidence from the peer-reviewed literature was sought out. The aim of the first pilot study was to select a pool of items that were common and mid-range believable, to allow for either reduction or increase in belief following retractions or affirmations, respectively. The second pilot study was run to ensure that the inference questions were in fact indirect measures of belief (i.e., that they correlated with the associated explicit belief measures; e.g., to ensure the inference question 'What percentage of lies can FBI detectives catch just by looking at physical tells' is sufficiently measuring an individual's belief that it is possible for liars to give themselves away by physical tells).

Pilot Study 1. The aim of the first pilot study was to select an item pool of myths that were common and at least mid-range believable. Thirty-one undergraduate students from the University of Western Australia took part. Participants indicated for 55 myth and 20 fact items (1) if they had heard of the item before (i.e., familiarity) and (2) the extent to which they believed each item (i.e., believability). Familiarity was measured on a five point scale ranging from "Definitely not heard before" to "Definitely heard before". Myths were removed from the stimulus set if they scored below a mean of 3.5 out of 5. Believability was measured on a 1-10 scale ranging from "Not at all believable" to "Very much so". Myths were removed from the stimulus set if they scored below a mean of 4.5; one additional item with a mean greater than 9.0 was also removed (to avoid any ceiling effects reducing the likelihood of a familiarity backfire effect).

After Pilot Study 1, there were 37 myths remaining. The mean familiarity score of the myths was $M = 4.46$ out of 5 ($SD = .35$). The mean familiarity rating of the facts was $M = 3.39$ ($SD = 1.04$). The believability mean for the myths was $M = 6.24$ ($SD = 1.12$) and the mean for the facts was $M = 5.34$ ($SD = 1.89$). Pre-manipulation familiarity and belief ratings were positively correlated, $r = .79$, showing that the more familiar items were believed more strongly.

Pilot Study 2. The second pilot study was run to ensure that the inference questions were in fact an indirect measure of belief in the initial claims. Participants were 100 individuals who volunteered via Crowdflower (http://www.crowdflower.com), a crowdsourcing website where contributors perform tasks and are paid for their services. Participants were paid $1.80. Five fact items were added to the set after Pilot Study 1 in order to boost their number in comparison to myths. Participants rated how much they believed in the 38 myths and 25 facts, and responded to two corresponding inference questions per claim.

Participants were excluded if they reported their English skills to be only "fair" (0 on a 4-point scale ranging from "fair" to "native speaker"; 5 individuals), if they took less than 15 minutes to complete the task (23 individuals) or more than 85 minutes (3 individuals; mean completion time was $M = 34.88$ minutes, $SD = 35.36$ minutes). The data were also screened for inconsistent response patterns suggestive of participants not paying attention, but no participants were excluded. A total of $N = 75$ participants were included in the analysis. Spearman's correlations were calculated for each item and the two corresponding inference questions. Items were excluded from the stimulus set if both inference questions did not significantly correlate with belief in the corresponding claim (with $p < .05$; r ranging from .23 – .81); this resulted in exclusion of 19 items—14 myths and 5 facts—leaving 24 myths and 20 facts.

In a final step, the four remaining myths with the lowest belief ratings and correlations between inference questions and belief ratings were removed. The final stimulus set thus comprised 20 myths and 20 facts, each with two corresponding inference questions.

Procedure. Participants were seated individually in testing booths and the experiment was administered by Qualtrics survey software. Participants were presented the 40 items in randomized order, and they indicated on a 0-10 scale the extent to which they believed each item using a computer mouse. Directly after each item was rated, participants received either a brief or a detailed explanation, which were randomly counter-balanced. In the immediate test condition (i.e., no retention interval), the test phase began immediately after all items had been rated and retracted or affirmed. The test phase involved a block of 80 inference questions (two per item) in random order, followed by a block of 40 direct belief ratings in random order. Participants in the 30-minute retention interval group completed an unrelated filler task before the test, and participants in the one-week group completed the test phase a week later—this test was administered in an online format in order to keep participation rates high. The test phase was identical regardless of retention interval.

Results

Belief ratings. Both pre-manipulation facts and myths attracted mid-range initial belief ratings, as expected, $M_{facts} = 5.69$, $SD_{facts} = .79$; $M_{myths} = 6.03$, $SD_{myths} = .97$. A within-subjects ANOVA comparing the pre-manipulation fact and myth belief ratings showed that participants initially believed the myths slightly more than the facts, $F(1,92) = 9.61$; $p = .003$; $MSE = .58$; $\eta_p^2 = .10$.

After participants read the affirmations/ corrections, participants' belief for facts increased, and belief for myths decreased, as shown in Figure 1. This belief change was

sustained temporarily for both myths and facts, yet after a one-week period belief for myths regressed. As post-manipulation belief levels remained below pre-manipulation belief levels, no true backfire effect was elicited.[2]

A 2 × 2 × 3 within-between ANOVA (with factors type of item, type of explanation, and retention interval) was performed on the post-manipulation belief ratings. For this and all further statistical analyses, belief ratings and inference scores for myths were reverse-coded. This was to simplify the analysis and allow the type of explanation (brief vs. detailed) to register as a main effect rather than an interaction. The figures and discussion of the data trends are presented in the original untransformed format to facilitate interpretation.

The analysis revealed three significant main effects. The main effect of type of item (myth vs. fact), $F(1,90) = 27.57$; $p < .001$; $MSE = 1.68$; $\eta_p^2 = .23$, indicated that overall fact belief ratings were closer to the ceiling than myth belief ratings were to the floor (Figure 1). The main effect of type of explanation (brief vs. detailed), $F(1,90) = 15.38$; $p < .001$; $MSE = .74$; $\eta_p^2 = .15$, indicated that detailed explanations were slightly better at eliciting belief change than brief explanations. The main effect of retention interval, $F(2,90) = 4.78$; $p = .011$; $MSE = 5.40$; $\eta_p^2 = .10$, indicated that the extent of belief change differed over time. This was qualified by a significant interaction of type of item and retention interval, $F(2,90) = 8.65$; $p < .001$; $MSE = 1.68$; $\eta_p^2 = .16$, indicating that the change in belief over time was different for facts and

[2] Nyhan et al. (2014) found that corrective information regarding the flu vaccine reduced participants' intent to vaccinate, but *only* in participants with high levels of concern about vaccine side effects. To address the assumption that backfire effects may only occur when correcting strong belief in the original misconception, the analysis was replicated using each participant's 30 % most strongly believed myths and 30 % least believed facts. There was no backfire effect observed for this subset of materials—i.e. myths that participants correctly assumed to be false, and facts that participants correctly assumed to be true. At one week, myth belief was not statistically different from pre-manipulation levels, $p > .05$.

myths, with fact belief remaining stable across intervals and myth belief rebounding over time(all other effects, $F < 1$).

Next, we ran a 2 × 2 × 2 within-between ANOVA (factors type of item, type of explanation, and retention interval) restricted to the 30-minute and one-week retention intervals to clarify specifically whether the difference between fact and (reverse-coded) myth ratings was greater after a week than 30 minutes, or in other words, whether belief change was more stable over time for myths versus facts. The interaction between type of item and retention interval was significant, $F(1,61) = 13.90$; $p < .001$; $MSE = 1.89$; $\eta_p^2 = .19$, indicating that belief ratings were stable for facts from 30 minutes to one-week, whereas belief ratings for myths increased during this time period.

Even on an individual level, the items showed a consistent pattern: the retracted myths were more likely to show regression towards their pre-manipulation levels, whereas beliefs in affirmed fact items were relatively sustained over time. Only one myth item showed a numerically larger belief rating a week after correction compared to pre-manipulation belief levels.[3]

Inference ratings. Even if participants were successfully discounting misinformation in the direct belief ratings, they could still be using misinformation in their reasoning. To address this question, we analyzed participants' mean inference scores. All inference scores were significantly correlated at the $p < .05$ level with the respective belief ratings: myth-brief, $r = .80$; myth-detailed, $r = .78$; fact-brief, $r = .70$; and fact-detailed, $r = .67$. This indicates that inference questions supplied a valid indirect measure of belief.

[3] This exception was 'cancer screening is greatly beneficial' in the brief explanation condition, which had a mean pre-manipulation belief rating of 5.04, which rose to 5.33 after one week.

A 2 × 2 × 3 within-between ANOVA was performed on the inference scores (with factors type of item, type of explanation, and retention interval). The results mimicked the pattern obtained with the post-manipulation belief scores, as Figure 2 illustrates. There were main effects of type of item, $F(1,90) = 6.08$; $p = .016$; $MSE = 2.01$; $\eta_p^2 = .06$, and type of explanation, $F(1,90) = 17.29$; $p < .001$; $MSE = .53$; $\eta_p^2 = .16$, as well as retention interval, $F(2,90) = 6.17$; $p = .003$; $MSE = 3.96$; $\eta_p^2 = .12$. There was an interaction between type of item and retention interval, $F(2,90) = 3.54$; $p = .033$; $MSE = 2.01$; $\eta_p^2 = .07$, indicating that the stability of scores across time differed for myths and facts. There was also a marginally significant interaction of type of explanation and retention interval, $F(2,90) = 2.44$; $p = .093$; $MSE = .53$; $\eta_p^2 = .05$, suggesting that detailed explanations are particularly beneficial over time (all other effects, $F < 1.72$, $p > .19$).

Analogous to the belief ratings analysis, a 2 × 2 × 2 within-between ANOVA (with factors type of item, type of explanation, and retention interval) was run testing specifically whether inference scores were less stable over time for myths versus facts in the 30-minute to one-week interval. The type of item by retention interval interaction was significant, $F(1,61) = 5.83$; $p = .019$; $MSE = 2.24$; $\eta_p^2 = .09$, demonstrating that inference scores increased over a one-week period for myths in comparison to facts.

Returning to the omnibus 2 × 2 × 3 analysis, there was also a marginal interaction between type of explanation and retention interval, $F(2,90) = 2.44$; $p = .09$; $MSE = .53$; $\eta_p^2 = .05$, suggesting that inference scores were more stable over time after detailed explanations compared to brief explanations. To corroborate this notion, an interaction contrast was run contrasting brief against detailed explanations and the pooled immediate and 30-minute intervals against the one-week interval (assigning lambda weights of 1, -1, 1, -1, to myth-brief, myth-detailed, fact-brief

and fact-detailed, and 1, 1, -2 to the immediate, 30-minute, and one-week retention intervals, respectively). The contrast was significant, $F(1,90) = 4.44$; $p = .038$; $MSE = .53$, indicating that a detailed explanation had its greatest benefit after a long delay. A detailed discussion of the Experiment 1 results will be deferred until the Experiment 2 data are presented.

Experiment 2

Experiment 1 showed that belief change was more sustained after fact affirmation compared to myth retraction. Experiment 2 was a conceptual replication of Experiment 1 but tested older adults. As we noted at the outset, it is possible that older adults are more strongly susceptible to the effects of familiarity, as older adults have less efficient strategic memory processes than young adults, whereas automatic processing is relatively age-invariant (Prull et al., 2006). While it is difficult to pinpoint the exact age at which recollection begins to decline, a study by Bender, Naveh-Benjamin, and Raz (2010) suggested a marked decline around the age of 40, and many studies investigating age-related differences in familiarity and recollection have used an older adult population with a mean age in the 60s (e.g. Aizpurua, Garcia-Bajos, & Migueles, 2009; Bastin & Van der Linden, 2003) or 70s (Anderson et al, 2008; Fernandes & Manios, 2012; Prull et al., 2006).

Method

Experiment 2 was identical to Experiment 1, with two changes: (1) it was conducted with an older adult population; (2) an additional three-week retention interval condition was added to maximize the chances of eliciting the familiarity backfire effect, given the temporal stability of familiarity in contrast to the temporal volatility of recollection.

Participants. Participants were 124 older adults over the age of 50, who volunteered after reading an ethically-approved information sheet. Participants were recruited by advertising through the University of Western Australia website, Western Australian radio, and flyers around Perth. Participants were paid A$15 for their participation. Participants were screened using the Montreal Cognitive Assessment (MoCA); thirteen participants were excluded as they scored below the normal range of 26 to 30 (Nasreddine et al., 2005). An additional two participants did not complete the task. Our final sample thus included $N = 109$ participants, with 39 males and 70 females between 50 and 87 years of age ($M = 64.37$, $SD = 8.91$).

Procedure. The procedure replicated Experiment 1, although prior to the study participants received the MoCA. One-week and three-week surveys were completed in an online format in order to keep participation rates high; two participants in the delayed conditions opted to receive paper copies of the survey. These were mailed back to the researchers once they had been completed.

Results

Belief ratings. A within-subjects ANOVA was performed on the pre-manipulation myth and fact belief ratings, which uncovered no significant differences between conditions, $M_{facts} = 6.10$, $SD_{facts} = 1.01$; $M_{myths} = 5.92$, $SD_{myths} = 1.04$. This indicates that prior to reading the explanations, participants believed myths and facts equally.

After participants read the explanations, the belief for facts increased and belief for myths declined, as can be seen in Figure 3. In striking similarity to Experiment 1, belief for facts was sustained over a one-week period, whereas belief for myths regressed between 30 minutes and one-week. Between week 1 and week 3, belief scores for both facts and myths regressed to a

similar extent. As post-manipulation myth belief levels remained below pre-manipulation belief levels, no true backfire effect was elicited.[4]

For all further analyses, belief ratings and inference score ratings for myths were reverse-coded, as in Experiment 1. A 2 × 2 × 4 within-between ANOVA on belief ratings was run, with within-subjects factors type of item (myth vs. fact) and type of explanation (veracity explained either briefly or in some detail), and the between-subjects factor retention interval (immediate, 30-minute, one-week, or three-weeks). The analysis revealed three significant main effects. The main effect of type of item (myth vs. fact), $F(1,105) = 30.39$; $p < .001$; $MSE = 3.04$; $\eta_p^2 = .0.22$, indicated that fact ratings were closer to the ceiling than myth ratings were to the floor. The main effect of type of explanation (brief vs. detailed), $F(1,105) = 14.91$; $p < .001$; $MSE = 1.08$; $\eta_p^2 = .12$, indicated that detailed explanations were better at eliciting belief change than brief explanations, and the main effect of retention interval, $F(3,105) = 11.56$; $p < .001$; $MSE = 5.36$; $\eta_p^2 = .25$, indicated that belief change differed over time. A significant interaction of type of item and retention interval, $F(3,105) = 4.37$; $p = .006$; $MSE = 3.04$; $\eta_p^2 = .11$, indicated that the change in belief over time was different for facts and myths, with fact belief remaining largely stable across intervals and myth belief increasing over time. The interaction of type of item and type of explanation was also significant, $F(1,105) = 4.75$; $p = .031$; $MSE = 1.11$; $\eta_p^2 = .04$, indicating that detailed explanations were slightly more effective for facts than for myths. Lastly, the interaction of type of explanation and retention interval indicated that detailed explanations promoted belief change better than brief explanations particularly after a long delay,

[4] To address the assumption that backfire effects may only occur when correcting strong belief in the original misconception, the analysis was replicated using each participants' 30 % most strongly believed myths and 30 % least believed facts. The trend in was replicated, and no backfire effect was elicited.

$F(3,105) = 3.83$; $p = .012$; $MSE = 4.11$; $\eta_p^2 = .10$. The remaining interaction of type of item, type of explanation, and retention interval remained non-significant, $F < 1$.

In subsequent contrast analyses, we focused first on the 30-minute and 1-week retention intervals (analogous to Experiment 1). An interaction contrast between type of item (myth vs. fact) and retention interval (30-minute vs. one-week) demonstrated that the belief difference between myths and facts was greater after one week than 30 minutes, $F(1,52) = 8.11$; $p = .006$; $MSE = 2.49$; $\eta_p^2 = .13$. Thus, fact belief remained stable over time whereas myth belief increased over the period of one week.

Focusing on retention intervals of 1 and 3 weeks, the analogous type of item by retention interval contrast was not significant, $p > .05$, while a contrast comparing week-1 and week-3 ratings collapsing all conditions across item and explanation levels was significant, $F(1,52) = 7.08$; $p = .010$; $MSE = 4.79$, indicating that from week 1 to week 3, fact and myth belief ratings regressed equivalently. In other words, item validity, in general, was being forgotten between a one and three-week period.

As post-manipulation myth belief significantly correlated at the $p < .05$ level with age for both brief retractions, $r = .21$, and detailed retractions, $r = .28$, a final set of belief-rating analyses looked at age at a finer level of granularity. Specifically, to further address the assumption that myth-belief updating deteriorates with age, a median-split analysis comparing participants aged 50-64 (“middle-aged” participants) with those 65 and older (“old” participants) was conducted (see Figure 4). Investigating type of explanation (brief vs. detailed), age (middle-aged vs. old), and retention interval (immediate vs. 30-minute vs. one-week), a 2 × 2 × 4 within-between ANOVA on myth beliefs was performed. This analysis yielded a main effect of age, $F(1,101) = 8.73$; $p = .004$; $MSE = 5.29$, $\eta_p^2 = .08$, indicating that old participants were less likely

to show sustained myth-belief change than middle-aged participants, and a main effect of retention interval $F(3,101) = 12.73$; $p < .001$; $MSE = 5.29$, $\eta_p^2 = .27$, indicating that belief changed over time (all other effects, $F < 2.77$, $p > .07$). Moreover, a 2 × 2 × 2 within-between ANOVA (with factors type of explanation, retention interval, and age) focusing on one-week and three-week retention intervals revealed a type of explanation by age interaction, $F(1,50) = 4.07$; $p = .049$; $MSE = 1.38$; $\eta_p^2 = .08$, indicating that after longer delays, detailed retractions led to more sustained belief change in comparison to brief retractions for middle-aged but not old participants.[5]

For the sake of completeness, a 2 × 2 × 4 within-between ANOVA on fact beliefs was performed (with factors type of item, type of explanation, and retention interval). The analysis yielded significant main effects of type of explanation, $F(1,101) = 19.20$; $p < .001$; $MSE = 1.03$; $\eta_p^2 = .16$, showing that detailed explanations were more effective than brief ones, age, $F(1,101) = 5.60$; $p = .020$; $MSE = 2.61$; $\eta_p^2 = .05$, indicating that old participants showed less belief change than middle-aged participants, and retention interval, $F(3,101) = 5.00$; $p = .003$; $MSE = 2.61$; $\eta_p^2 = .13$, showing that belief changed over time (all other effects, $F < 1.23$, $p > .30$). In an analysis confined to the immediate, 30-minute, and one-week conditions, the retention

[5] A 2 × 2 × 3 ANOVA also including the young adults from Experiment 1 (with factors type of explanation [brief and detailed], retention interval [30 minutes and one-week], and age [young adults, middle aged, and old], on post-explanation myth scores likewise revealed a main effect of age, $F(1,111) = 3.43$; $p = .036$; $MSE = 4.87$; $\eta_p^2 = .06$. An interaction contrast revealed that young adults and middle aged participants were equivalently better at reducing their belief in misconceptions than older adults, $p = .015$. Unlike the above analysis, there is no interaction of explanation and age—presumably because we could not include the week 3 ratings in the analysis—however, a planned comparison contrasting old adults against pooled young and middle aged adults, and brief explanations against detailed explanations, approached significance, $F(1,111) = 3.03$; $p = .084$; $MSE = 1.15$.

interval effect was non-significant, $p > .05$, demonstrating that for the duration of one week, fact belief was sustained.

Inference ratings. Returning to the analysis of the full sample, inference scores are presented in Figure 5. A 2 × 2 × 4 within-between ANOVA (with factors type of item, type of explanation, and retention interval) on the inference scores revealed main effects of type of item, $F(1,105) = 13.47$; $p < .001$; $MSE = 2.00$; $\eta_p^2 = .11$, type of explanation, $F(1,105) = 14.89$; $p < .001$; $MSE = .85$; $\eta_p^2 = .12$, and retention interval, $F(3,105) = 6.43$; $p < .001$; $MSE = 4.24$; $\eta_p^2 = .16$., as well as an interaction between type of item and retention interval, $F(3,105) = 3.75$; $p = .013$; $MSE = 2.00$; $\eta_p^2 = .10$, suggesting that the stability of scores across time differed for facts and myths. An interaction contrast, analogous to Experiment 1, testing whether inference scores were more stable for facts versus myths in the 30-minute to one-week interval, was significant, $F(1,52) = 9.01$; $p = .004$; $MSE = 1.74$; $\eta_p^2 = .15$.

There was also an interaction of type of explanation and retention interval $F(3,105) = 3.38$; $p = .021$; $MSE = .85$; $\eta_p^2 = .09$, indicating that inference scores were more stable across time after detailed explanations compared to brief explanations. To corroborate this notion, an interaction contrast was run contrasting brief against detailed explanations and the pooled immediate and 30-minute intervals against the pooled one-week and three-week intervals. The contrast was significant, $F(1,105) = 10.07$; $p = .002$; $MSE = .85$, indicating that a detailed explanation had its greatest benefit after a long delay (all other effects, $F < 1.30$, $p > .25$).

Discussion

The present research aimed to determine the parameters of differential forgetting of myth and fact veracity over time, in order to clarify if and under what conditions familiarity may contribute to false acceptance of corrected myths as true. Dual-process accounts of continued

influence effects of misinformation (e.g., Ecker et al., 2010) suggest that post-correction reliance on misinformation can be based on automatic memory processes (i.e., myth familiarity) in the absence of strategic retrieval and control processes. Hence familiarity-based acceptance of corrected falsehoods could be a mechanism underlying continued influence effects of misinformation. To investigate this, we presented participants with both myths and facts, obtained a pre-manipulation belief rating, then corrected the former and affirmed the latter. We manipulated factors known to affect strategic memory processes, thus varying the relative impact of familiarity. Specifically, we manipulated the explanations' level of detail and retention interval, and contrasted age groups, and we measured how these factors affected people's post-explanation beliefs and inferences.

While some studies have shown a continued influence effect after a brief retention interval (e.g. Ecker et al., 2011; Johnson & Seifert, 1994), our corrections (and affirmations) were found to be relatively effective in the short-term. This difference may be due to the fact that, unlike the typical continued-influence paradigm, we retracted simple statements rather than causal relationships regarding an event, which may be particularly resistant to correction. The short-term efficacy of the explanations was more apparent for direct belief ratings (e.g., see Figure 1), whereas our inference measure (e.g., see Figure 2) closely resembled the typical result pattern found in continued-influence studies, which often also use inference questions to assess misinformation effects.

Differential Forgetting of Myths and Facts Over Time

Across both experiments, we found a striking asymmetry in that belief change was more sustained after fact affirmation compared to myth retraction—retractions thus seemingly have an "expiration date". This asymmetry could be partially explained by familiarity. In the case of an

affirmed fact, it does not matter if an individual relies on the recollection of the affirmation or on the boosted familiarity of the factual statement—familiarity and recollection operate in unison and lead to the individual assuming the item to be true. However, in the case of a retracted myth, recollection of the retraction will support the statement's correct rejection, whereas the myth's boosted familiarity will foster its false acceptance as true, as familiarity and recollection stand in opposition (Jacoby, 1991).

Our inference results mirrored the trend obtained with the belief ratings, demonstrating that familiarity effects can extend to inferential reasoning and potentially decision making. It is even possible that the act of responding to inference questions can contribute to increased familiarity of the misconception, in that the information is subjectively re-experienced during memory retrieval following exposure to the inference question, once again leading to a potentially increased perception of validity (Ozubko & Fugelsang, 2011).

Age and Level of Detail

Overall, the pattern of belief change over time—and in particular the asymmetry between facts and myths—was similar in young and older participants. Even young adults' recollection fades over time, leading to an increased reliance upon familiarity in judging the veracity of information (Gilbert et al., 1990). However, "old" participants aged 65 and over were found to be comparatively worse than those aged 50-64 ("middle-aged" participants) at sustaining their post-correction belief that myths are inaccurate. This supports the notion that older adults have less efficient strategic memory processes and thus less effective retrieval of the link between an item and contextual details (Naveh-Benjamin, 2000; Prull et al., 2006). As the mnemonic link between a statement and its veracity is weaker in older adults (Glisky et al., 2001), they seem particularly susceptible to the "re-believing" of myths. Although there was also a significant

difference in fact belief between the "middle-aged" and "old" groups, this reflected the fact that the old participants were less likely to initially update their belief immediately after the affirmation. This differed from myth belief where belief change immediately after a correction was substantial yet followed by relatively steep forgetting as time progressed.

Detailed refutations seemed to somewhat mitigate the negative impact of familiarity in both younger and middle-aged adults. This is supported by parts of the educational literature, which highlight the benefits of detailed refutations (Tippett, 2010). Refutations may encourage participants to detect inconsistencies between their own inaccurate beliefs and the corrective information, leading to a facilitation of belief change even over long delays (Bedford & J. Cook, 2013; Guzzetti, 2000; Kowalski & Taylor, 2009). The benefit of directly addressing misconceptions could additionally be explained by detailed explanations fostering skepticism regarding the initial misinformation or its source (cf. Lewandowsky, Stritzke, Freund, Oberauer, & Krueger, 2013; Lewandowsky, Stritzke, Oberauer, & Morales, 2005). However, as much of this research stems from the educational literature, it has mostly used undergraduates or school-age participants (Guzzetti et al., 1993). The current study found that for "old" adults over the age of 65, correcting myths using detailed refutations was as ineffective as brief retractions.

The Familiarity Backfire Effect

The present research provides evidence for familiarity causing an increase in post-correction myth belief after a delay; this meshes well with previous studies that similarly reported that myths are often "misremembered" as facts over time (Peter & Koch, 2016; Skurnik et al., 2005; Skurnik et al., 2007). However, we found no evidence for the existence of a true familiarity-based *backfire* effect. As in these previous studies, the corrections did help participants update their beliefs in the right direction—that is, myth beliefs were reduced by the

corrections. Corrections repeating the myth were simply less effective (compared to fact affirmations) rather than backfiring.

The lack of a familiarity backfire effect conforms to a range of theoretical proposals which suggest that repeating misinformation when correcting could even *facilitate* belief updating. Stadtler, Scharrer, Brummernhenrich, and Bromme (2013) as well as Putnam, Wahlheim, and Jacoby (2014) proposed that the detection of conflict—which is arguably made more salient through repetition of the misinformation during its retraction—is beneficial for updating. Reconsolidation theory likewise argues that reminders of to-be-corrected information will labilize its memory representation, thereby facilitating updating (Hardt, Einarsson, & Nader, 2010). Finally, Kendeou, Walsh, Smith, and O'Brien, (2014) argued that outdated and new information must be co-activated for knowledge revision to occur. This is consistent with a study by Pashler, Kang, and Mozer (2013), who found that repeating the original misinformation prior to learning new information enhanced memory for the new information when tested one week later.

This implies that future research still faces a conundrum: while the present findings suggest that false acceptance of corrected myths as true is at least partially driven by familiarity, it seems that corrections that do not repeat the myth may be even less effective than corrections that do repeat the myth (e.g., Ecker, Hogan, & Lewandowsky, in press; Wilkes & Leatherbarrow, 1988). In other words, if a myth is not repeated when corrected, the associated lack of salience, conflict detection, and/or myth/correction co-activation may be even more detrimental to belief updating than the boost of the myth's familiarity.

Potential Limitations and Future Directions

Obtaining belief measures prior to the experimental manipulation could be considered a limitation as it may have influenced how the corrective explanations were processed. However, in our opinion it is likely that a person's belief is spontaneously cued when a statement of unclear veracity (e.g., a potentially dubious news headline) is encountered, or when a correction is presented by itself (e.g., if one is told that listening to Mozart does not increase IQ, it seems likely that one would consider whether or not one believes the original claim). Thus, asking for an explicit expression of belief prior to a correction will not necessarily have a strong impact on how the correction is processed. In our view, from a methodological perspective, the advantages of a pretest-posttest design outweigh the disadvantages. "Posttest-only with control" designs as used by Skurnik et al. (2005)—where one group received the correction and another group received no correction—can be considered quasi-experimental as the treatment and control groups cannot be adequately compared at baseline (Morris, 2008). This potentially reduces internal validity because the differences at posttest may be artificially inflated (T. D. Cook & Campbell, 1979; Morris & DeShon, 2002).

The artificial nature of the task could be seen as another limitation, as participants evaluated a long series of statements. However, people often process a large number of news headlines in a short period of time (e.g., when skimming a newspaper or scanning one's social media feed), arguably assessing or at least monitoring the truth/belief status of each. Thus, we argue that people routinely deliberate belief prior to correction (i.e., in an experimental context, before a post-correction belief rating), even with large numbers of statements.

We have interpreted our finding that myths are more likely than facts to be misremembered after a delay as an effect of familiarity when strategic memory is limited. The

present research focused on factors that influence strategic memory processes; future research could test the proposed relationship between familiarity and misinformation effects more directly, for example by correcting statements that are familiar to some participants but not others. Previous research has found that misinformation effects are particularly strong if the misinformation is repeatedly presented before a correction (Ecker et al., 2011, also see Weaver, Garcia, Schwarz, & Miller, 2007), in line with the familiarity notion.

Moreover, future research could apply alternative testing procedures to further investigate the mechanisms underlying the effects reported here. For example, if myth acceptance is familiarity-driven, one might expect corrected myths to be accepted as true particularly in tasks requiring true/false categorization of statements (which may be more recognition-based) rather than in tasks that have a stronger recall component.

Practical Applications

The applied goal of this research was to provide empirically-based advice on how to correct misconceptions. The present data suggest the following: First, corrections should include details as to why the misinformation is incorrect, as detailed refutations are more effective than brief retractions, particularly with younger participants. Thus the misinformation should be explicitly retracted and paired with a comprehensive rebuttal.

Second, even the efficacy of detailed refutations of familiar misconceptions will lessen over time, and important corrections may need to be provided repeatedly, despite the potential risks of further boosting the myth's familiarity (also see Ecker et al., 2011a). While this recommendation seems somewhat ironic in the context of the boosted-familiarity notion, boosting the more volatile recollection of the correction to offset myth familiarity may be necessary to achieve enduring belief change.

Third, explicitly mentioning a familiar misconception within a retraction will not typically backfire in the true sense of the word (this qualifies earlier recommendations; e.g., J. Cook & Lewandowsky, 2011; Lewandowsky et al., 2012). Repeating the myth when retracting it may be crucial for belief updating because it increases the correction's salience and fosters conflict detection and co-activation of myth and correction (Kendeou et al., 2014; Putnam et al., 2014; Stadtler et al., 2013). However, given the aforementioned trade-off between the harm from boosting myth-familiarity and the benefit from boosting recollection of the correction (e.g. the association of the myth and its "negation-tag"), theoretically there may be circumstances where the harm outweighs the benefit. Moreover, it may also be problematic to circulate corrections if individuals have not previously encountered the relevant misconception, as this may potentially make the misinformation familiar to new audiences (Schwarz et al., 2016). It follows that, after correcting a myth, the focus should be placed upon factual information as much as possible in order to avoid boosting myth familiarity more than necessary (cf. Ecker et al., 2010; Johnson & Seifert, 1994; Lewandowsky et al., 2012; Seifert, 2002).

References

Aizpurua, A., Garcia-Bajos, E., & Migueles, M. (2009). False memories for a robbery in young and older adults. *Applied Cognitive Psychology, 23*, 174–187.

Anderson, N. D., Ebert, P. L., Jennings, J. M., Grady, C. L., Cabeza, R., & Graham, S. J. (2008). Recollection- and familiarity-based memory in healthy aging and amnestic mild cognitive impairment. *Neuropsychology, 22*, 177–187.

Ayers, M. S., & Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin & Review*, *5*, 1–21.

Bastin, C., & Van der Linden, M. (2003). The contribution of recollection and familiarity to recognition memory: A study of the effects of test format and aging. *Neuropsychology, 17*, 14–24.

Begg, I..M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General, 121*, 446-458.

Bender, A. R., Naveh-Benjamin, M., & Raz, N. (2010). Associative deficit in recognition memory in a lifespan sample of healthy adults. *Psychology and Aging*, *25*, 940–948. http://doi.org/10.1037/a0020595

Berinsky, A. J. (2015) Rumors and health care reform: Experiments in political misinformation. *British Journal of Political Science*, 1-22. doi:10.1017/S0007123415000186

Brown, M. W., & Warburton, E. C. (2006). Associations and dissociations in recognition

memory systems. In H. Zimmer, A. Mecklinger & U. Lindenberger (eds.), *Handbook of binding and memory: Perspectives from cognitive neuroscience* (pp. 342-432) Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198529675.003.0016

Brainerd, C. J., & Reyna, V. F. (2008). Episodic over-distribution: A signature effect of familiarity without recollection. *Journal of Memory and Language, 58*, 765–786. http://doi.org/10.1016/j.jml.2007.08.006

Cameron, K. A., Roloff, M. E., Friesema, E. M., Brown, T., Jovanovic, B. D., Hauber, S., & Baker, D. W. (2013). Patient knowledge and recall of health information following exposure to "facts and myths" message format variations. *Patient Education and Counseling*, *92*, 381–387. http://doi.org/10.1016/j.pec.2013.06.017

Campbell, D. (1997) *The Mozart Effect: Tapping the power of music to heal the body, strengthen the mind, and unlock the creative spirit.* New York, NY: Avon Books Inc.

Cook, J., Bedford, D., & Mandia, S. (2014). Raising climate literacy through addressing misinformation: Case studies in agnotology-based learning. *Journal of Geoscience Education*, *62*, 296–306.

Cook, J., & Lewandowsky, S. (2011). *The debunking handbook*. Retrieved from http://www.skepticalscience.com/docs/Debunking_Handbook.pdf

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Boston: Houghton Mifflin.

Connolly, K., Chrisafis, A., McPherson, P., Kirchgaessner, S., Haas, B., Phillips, D., Hunt, E., & Safi, M. (2016). Fake news: an insidious trend that's fast becoming a global problem. *The Guardian.* Available from https://www.theguardian.com/media/2016/dec/02/fake-news-facebook-us-election-around-the-world

Diakidoy, I. A. N., Mouskounti, T., Fella, A., & Ioannides, C. (2016). Comprehension processes and outcomes with refutation and expository texts and their contribution to learning. *Learning and Instruction, 41*, 60–69. http://doi.org/10.1016/j.learninstruc.2015.10.002

DiFonzo, N., Beckstead, J. W., Stupak, N., & Walders, K. (2016). Validity judgments of rumors heard multiple times: The shape of the truth effect. *Social Influence, 11*, 22–39.

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends in Cognitive Sciences, 11*, 379–386.

Dechene, A., Stahl, C., Hansen, J., & Wanke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review, 14*, 238–257. http://doi.org/10.1177/1088868309352251

Eagly, A. H. & Chaiken, S. (1993). *The psychology of attitudes.* Florida: Harcourt Brace Jovanovich College Publishers.

Ecker, U. K. H., Hogan, J. L., & Lewandowsky, S. (in press). Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition.*

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 & Review*, *18*, 570–578. http://doi.org/10.3758/s13423-011-0065-1

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

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*, *39*, 175–191.

Fernandes, M. A., & Manios, M. (2012). How does encoding context affect memory in younger and older adults? *The Quarterly Journal of Experimental Psychology*, *65*, 1699–1720.

Gemberling, T. M., & Cramer, R. J. (2014). Expert testimony on sensitive myth-ridden topics: Ethics and recommendations for psychological professionals. *Professional Psychology: Research and Practice*, *45*, 120–127.

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,* 601–613.

Glisky, E. L., Rubin, S. R., & Davidson, P. S. (2001). Source memory in older adults: An encoding or retrieval problem? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *27*, 1131–1146.

Guillory, J. J., & Geraci, L. (2013). Correcting erroneous inferences in memory: The role of source credibility. *Journal of Applied Research in Memory and Cognition*, *2*, 201–209. http://doi.org/10.1016/j.jarmac.2013.10.001

Guzzetti, B. J. (2000). Learning counter-intuitive science concepts: What have we learned from over a decade of research? *Reading & Writing Quarterly*, *16*, 89–98. http://doi.org/10.1080/105735600277971

Guzzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Promoting conceptual change in science: A comparative meta-analysis of instructional interventions from reading education and science education. *Reading Research Quarterly*, *28*, 117–159. http://doi.org/10.2307/747886

Hardt, O., Einarsson, E. Ö., & Nader, K. (2010). A bridge over troubled water: Reconsolidation as a link between cognitive and neuroscientific memory research traditions. *Annual Review of Psychology, 61*, 141–167. http://doi.org/10.1146/annurev.psych.093008.100455

Hart, P. S., & Nisbet, E. C. (2012). Boomerang effects in science communication. *Communication Research, 39,* 701–723. http://doi.org/10.1177/0093650211416646

Herron, J. E., & Rugg, M. D. (2003). Strategic influences on recollection in the exclusion task: Electrophysiological evidence. *Psychonomic Bulletin & Review*, *10*, 703–710.

Hoaglin, D. C., & Iglewicz, B. (1987). Fine-tuning some resistant rules for outlier labeling. *Journal of the American Statistical Association*, *82,* 1147–1149. http://doi.org/10.2307/2289392

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings.* Thousand Oaks, CA: Sage.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541. http://doi.org/10.1016/0749-596X(91)90025-F

Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1420–1436. http://doi.org/10.1037/0278-7393.20.6.1420

Kendeou, P., Walsh, E. K., Smith, E. R., & O'Brien, E. J. (2014). Knowledge revision processes in refutation texts. *Discourse Processes*, *51*, 374–397.

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language, 37,* 555–583.

Knowlton, B. J., & Squire, L. R. (1995). Remembering and knowing: Two different expressions of declarative memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 699–710.

Kowalski, P., & Taylor, A. K. (2009). The effect of refuting misconceptions in the introductory psychology class. *Teaching of Psychology*, *36*, 153–159. http://doi.org/10.1080/00986280902959986

Lavoipierre, A. (2017). 'Fake news' named 2016 Word of the Year by Macquarie Dictionary. *ABC News.* Available from http://www.abc.net.au/news/2017-01-25/fake-news-named-2016-word-of-the-year/8211056

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

Lewandowsky, S., Stritzke, W. G. K., Freund, A. M., Oberauer, K., & Krueger, J. I. (2013). Misinformation, disinformation, and violent conflict: From Iraq and the "War on Terror" to future threats to peace. *American Psychologist*, *68*, 487–501. http://doi.org/10.1037/a0034515

Lewandowsky, S., Stritzke, W. G. K., Oberauer, K., & Morales, M. (2005). Memory for fact, fiction, and misinformation: The Iraq War 2003. *Psychological Science*, *16*, 190–195.

Lilienfeld, S. O., Marshall, J., Todd, J. T., & Shane, H. C. (2014). The persistence of fad interventions in the face of negative scientific evidence: Facilitated communication for

autism as a case example. *Evidence-Based Communication Assessment and Intervention, 8,* 62–101.

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods, 11,* 364–386.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105–125.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., … Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*, 695–699.

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1170–1187. http://doi.org/10.1037//0278-7393.26.5.1170

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*, 303–330. http://doi.org/10.1007/s11109-010-9112-2

Nyhan, B., & Reifler, J. (2015). Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. *Vaccine*, *33*, 459–464. http://doi.org/10.1016/j.vaccine.2014.11.017

Nyhan, B., Reifler, J., Richey, S., & Freed, G. L. (2014). Effective messages in vaccine promotion: A randomized trial. *Pediatrics*, *133*, e835–e842. http://doi.org/10.1542/peds.2013-2365

Ozubko, J. D., & Fugelsang, J. (2011). Remembering makes evidence compelling: Retrieval from memory can give rise to the illusion of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 270–276. http://doi.org/10.1037/a0021323

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,* 3–25.

Pietschnig, J., Voracek, M., & Formann, A. K. (2010). Mozart effect - Shmozart effect: A meta-analysis. *Intelligence, 38,* 314-323.

Prull, M. W., Dawes, L. L. C., Martin, A. M., Rosenberg, H. F., & Light, L. L. (2006). Recollection and familiarity in recognition memory: Adult age differences and neuropsychological test correlates. *Psychology and Aging, 21,* 107–118. http://doi.org/10.1037/0882-7974.21.1.107

Putnam, A. L., Wahlheim, C. N., & Jacoby, L. L. (2014). Memory for flip-flopping: Detection and recollection of political contradictions. *Memory & Cognition, 42,* 1198–1210. http://doi.org/10.3758/s13421-014-0419-9

Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences, 9,* 95–123.

Rauscher, F. H., Shaw, G. L., & Ky, C. N. (1993). Music and spatial task performance. *Nature, 365,* 611.

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8,* 385–407.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences, 11,* 251–257.

Seifert, C. M. (2002) The continued influence of misinformation in memory: What makes a correction effective? *Psychology of Learning and Motivation, 44,* 265–292.

Schwarz, N., Newman, E., & Leach, W. (2016). Making the truth stick and the myths fade: Lessons from cognitive psychology. *Behavioural Science and Policy, 2,* 85-95.

Schwarz, N., Sanna, L. J., Skurnik, I., & Yoon, C. (2007). Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information campaigns. *Advances in Experimental Social Psychology*, *39,* 127–161.

Silverman, C. (2016). Fake news expert on how false stories spread and why people believe them. *NPR.* Available from http://www.npr.org/2016/12/14/505547295/fake-news-expert-on-how-false-stories-spread-and-why-people-believe-them

Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research*, *31,* 713–724.

Skurnik, I., Yoon, C., & Schwarz, N. (2007). *Education about flu can reduce intentions to get a vaccination*. Unpublished manuscript.

Stadtler, M., Scharrer, L., Brummernhenrich, B., & Bromme, R. (2013). Dealing with uncertainty: Readers' memory for and use of conflicting information from science texts as function of presentation format and source expertise. *Cognition & Instruction*, *31,* 130-150.

Schultz, D. D. E. P. (2012). *Truth goggles: Automatic incorporation of context and primary source for a critical media experience.* Unpublished doctoral thesis, Massachusetts Institute of Technology. Retrieved from http://dspace.mit.edu/handle/1721.1/76530

Tippett, C. D. (2010). Refutation text in science education: A review of two decades of research. *International Journal of Science & Mathematics Education*, *8*, 951–970. http://doi.org/10.1007/s10763-010-9203-x

Toth, J. P. (1996). Conceptual automaticity in recognition memory: Levels-of-processing effects on familiarity. *Canadian Journal of Experimental Psychology*, *50*, 123–38.

Trevors, G. J., Muis,K. R., Pekrun, R., Sinatra, G. M., & Winne,P. H. (2016). Identity and epistemic emotions during knowledge revision: A potential account for the backfire effect. *Discourse Processes*, *53*, 339–370.

Wang, W.-C., Brashier, N. M., Wing, E. A., Marsh, E. J., & Cabeza, R. (2016). On known unknowns: Fluency and the neural mechanisms of illusory truth. *Journal of Cognitive Neuroscience*, *28*, 739–746.

Wilkes, A. L., & Leatherbarrow, M. (1988). Editing episodic memory following the identification of error. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *52A*, 165–183.

Weaver, K., Garcia, S.M., Schwarz, N., & Miller, D.T. (2007). Inferring the population of an opinion from its familiarity: A repetitive voice can sound like a chorus. *Journal of Personality and Social Psychology*, *92, 821-833.*

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

Yonelinas, A. P., & Jacoby, L. L. (2012). The process-dissociation approach two decades later: Convergence, boundary conditions, and new directions. *Memory & Cognition*, *40*, 663–680. http://doi.org/10.3758/s13421-012-0205-5

Zimmer, H. D., & Ecker, U. K. H. (2010). Remembering perceptual features unequally bound in object and episodic tokens: Neural mechanisms and their electrophysiological correlates. *Neuroscience & Biobehavioral Reviews*, *34*, 1066–1079. http://doi.org/10.1016/j.neubiorev.2010.01.014

Table 1

Example of a myth and fact, corresponding explanations, and inference questions

Myth	Liars sometimes give themselves away by physical 'tells' such as looking to the right or not looking you in the eye.
Brief retraction	Liars sometimes give themselves away by physical 'tells' such as looking to the right or not looking you in the eye. **MYTH** Liars do not give themselves away by physical 'tells' such as looking to the right or not looking you in the eye.
Detailed refutation	Liars sometimes give themselves away by physical 'tells' such as looking to the right or not looking you in the eye. **MYTH** Physical signals which are often assumed to be the 'tells' of a liar, are in fact signs of emotional discomfort in general. When a person is being interviewed or is accused of a crime, a non-liar is equally likely to express these signals. A meta-analysis of over 100 studies found no consistent physical cues when a person was lying. The experimenters stated that "there are no behaviors that always occur when people are lying and never occur when they are telling the truth".
Myth inference question	What percentage of lies can FBI detectives catch just by looking at physical tells? (0-100%)
Fact	Humans can regrow the tips of fingers and toes after they have been amputated.
Brief affirmation	Humans can regrow the tips of fingers and toes after they have been amputated. **FACT** Humans can regrow the tips of fingers and toes after they have been amputated.
Detailed affirmation	Humans can regrow the tips of fingers and toes after they have been amputated. **FACT** Astonishingly, humans have a very amphibian-like trait of being able to regenerate. Unfortunately, this is limited to the very tips of our fingers and toes. A study in 1970 found that if the individual was under the age of 10, they had a limited capability to even regrow bone. There are stem cells at the base of each nail, which aid ordinary nail growth as well as the ability to rebuild the digit tip after amputation. Interestingly, a regenerated finger will sometimes lack a fingerprint.
Fact inference question	What proportion of fingers will regenerate after the tip has been amputated? **(0-100 %)**

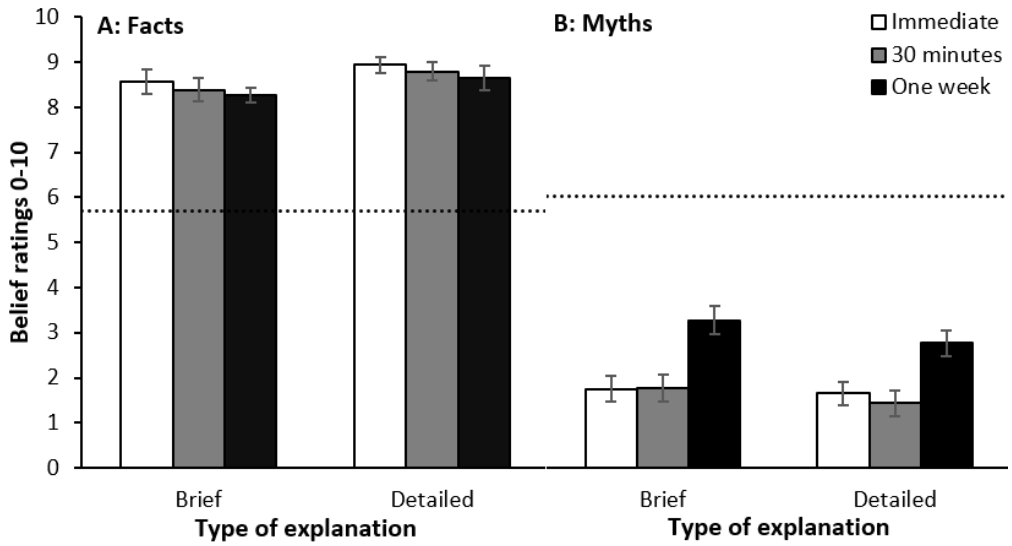

Figure 1. Post-manipulation belief ratings over time in Experiment 1. Dotted lines indicate the pre-manipulation belief ratings' mean.

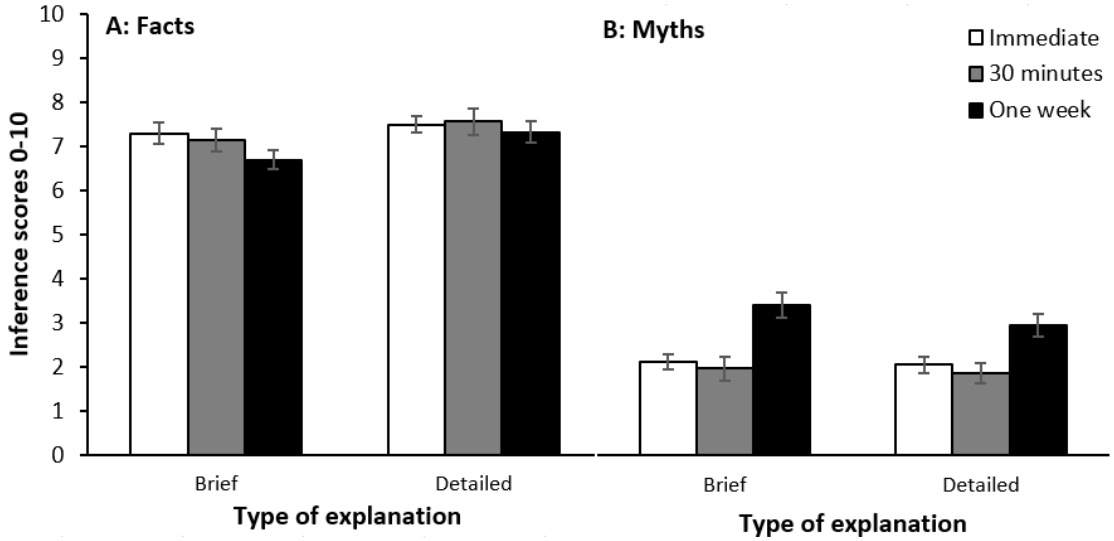

Figure 2. Post-manipulation inference scores in Experiment 1 over time.

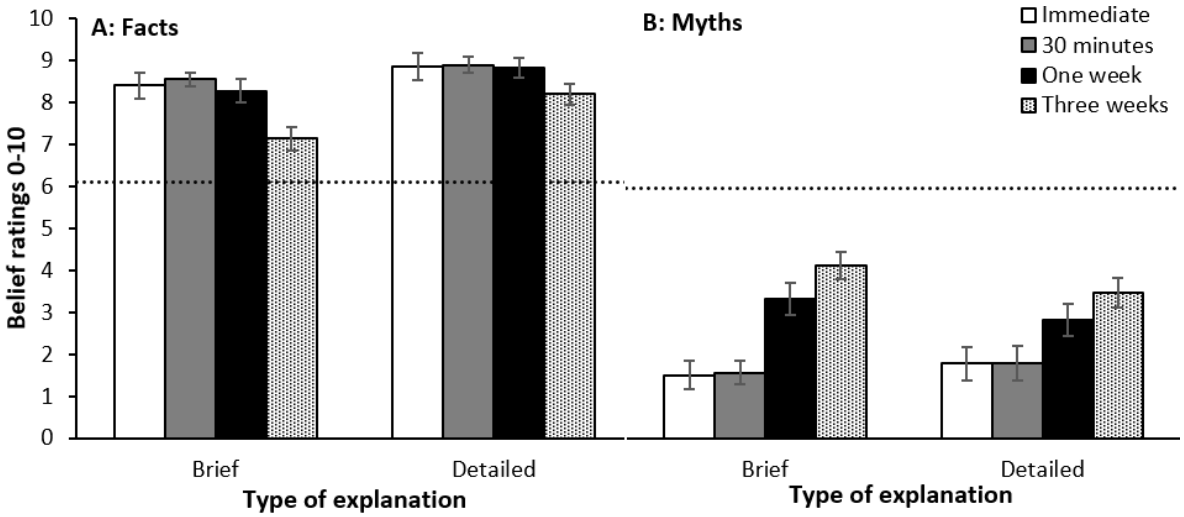

Figure 3. Post-manipulation belief ratings in older adults in Experiment 2. Dotted lines indicate the pre-manipulation belief ratings' mean.

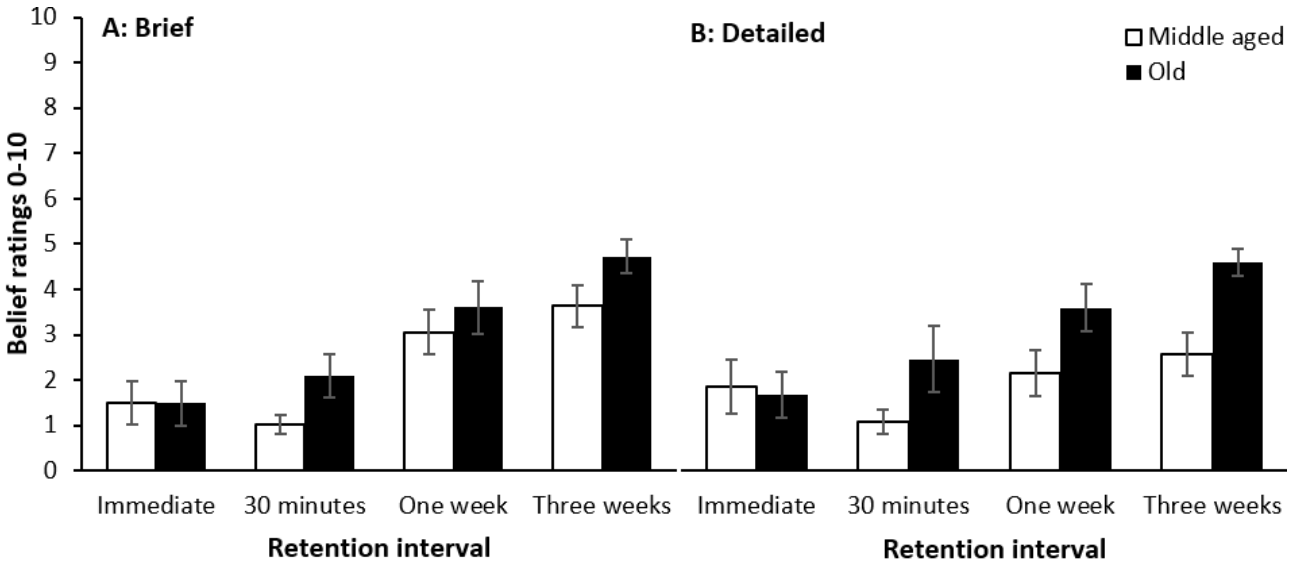

Figure 4. Post-manipulation myth belief ratings presented in an age-based median split in the older adult sample.

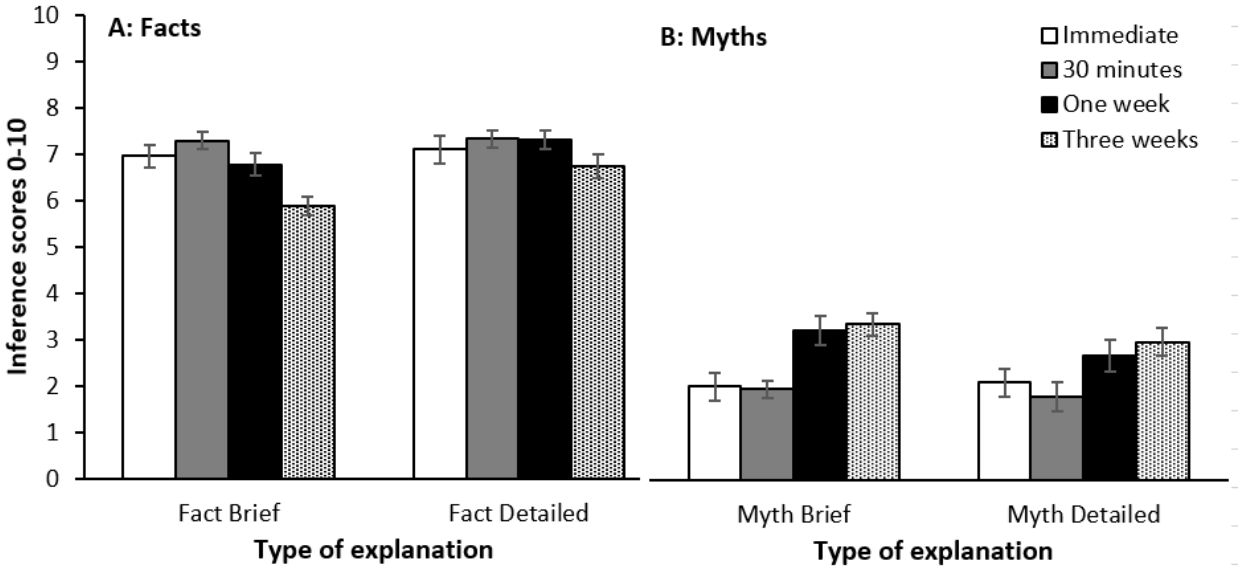

Figure 5. Post-manipulation inference scores in an older adult population in Experiment 2.