KEMP, PAIGE. Ph.D. Understanding the Effects of Fake News Corrections on Memory and Belief Accuracy: The Roles of Reminders, Retrieval, and Repetition. (2024)
Directed by Dr. Christopher Wahlheim. 232 pp.

Exposure to fake news can have detrimental effects on memories and beliefs, carrying widespread consequences for individuals and society. When it comes to correction strategies, there is an ongoing debate on whether corrections should repeat the fake news details or not. The familiarity backfire account argues that repeating fake news increases its familiarity and perceived accuracy, thereby impairing memory and belief accuracy. Conversely, integration-encoding accounts propose that repeating fake news can facilitate memory and belief accuracy by facilitating conflict detection and allowing both representations to be integrated together. In this integrated dissertation, three empirical papers are presented to build on previous work by exploring how fake news reminders, retrieval, and repeated exposure influences correction efficacy on memory and belief accuracy. The findings observed here more closely align with integration-encoding accounts than the familiarity backfire account, in showing that increasing accessibility to fake news can improve correction efficacy. However, it was also found that increasing accessibility to fake news can also impair memory and belief accuracy when corrections are not remembered, thus emphasizing the moderating role of recollection-based retrieval. Theoretical and practical implications of this work are discussed along with directions for future work to establish a more comprehensive understanding of how to effectively correct fake news.

UNDERSTANDING THE EFFECTS OF FAKE NEWS CORRECTIONS ON MEMORY AND BELIEF ACCURACY: THE ROLES OF REMINDERS, RETRIEVAL, AND REPETITION

by

Paige Kemp

A Dissertation
Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro

2024

Approved by

Dr. Christopher Wahlheim
Committee Chair

DEDICATION

I dedicate this dissertation to my parents, who have been instrumental in shaping both my academic journey and personal growth. Your love, encouragement, and unwavering belief in my potential have been the driving force behind this achievement.

To my dad, I want to thank you for always pushing me to do my best, whether it was in sports or academics. Thank you for being my constant support and for cheering me on every step of the way. To my mum, your positivity and support have meant the world to me, especially during tough times. Thank you for teaching me to find beauty in life's simple joys and for always seeing the world through hopeful eyes. I hope I have made you both proud.

APPROVAL PAGE

This dissertation written by Paige Kemp has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair ______________________
Dr. Christopher Wahlheim

Committee Members ______________________
Dr. Michael Kane

____
Dr. Peter Delaney

____
Dr. Blair Wisco

February 23, 2024
Date of Acceptance by Committee

February 23, 2024
Date of Final Oral Examination

ACKNOWLEDGEMENTS

I would like to acknowledge my advisor, Dr. Chris Wahlheim, for his mentorship throughout my graduate career. Under his guidance, I developed the skills necessary to become a productive and precise researcher, which were instrumental in completing this integrated dissertation. I am also deeply thankful to my committee members, Drs. Michael Kane, Peter Delaney, and Blair Wisco, for their invaluable contributions at both the preliminary and dissertation stages. Your insights and feedback have greatly enriched my work.

I also want to express my genuine appreciation to my friends Sydney Garlitch, AJ Garlitch, and Hayley Liebenow for their consistent support and invaluable assistance throughout my graduate career. They've been my rock every step of the way, whether it's been lending an ear during tough academic times or celebrating my accomplishments. Their friendship means the world to me, I know we will be lifelong friends. Thank you for everything!

Finally, I also want to give a shoutout to my cat, Roo, for keeping me company and bringing a little extra warmth to my late-night writing sessions. Thanks for being the most perfect little cat a graduate student could ask for! Lastly, I want to express my deepest gratitude to my lovely family—my nan, Evelyn Kemp, my brother, Jordan Kemp, my sister-in-law, Ellie-Rose Kemp, my nephew, Albie Kemp, and niece, Edith-Rose Kemp. Your unconditional love, laughter, and support have been a constant source of encouragement and inspiration. I am grateful for the love and care you have provided me throughout this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTEGRATED INTRODUCTION

We live in a post-truth world where we are constantly sifting through what's real and fake – it is like a never-ending game of fact-checking. Although fake news isn't an entirely novel phenomenon, the surge in interest surrounding the prevalence and impact of fake news has increased since the events of the US 2020 Election and Brexit (Vosoughi et al., 2018; Höller, 2021). The fundamental concern revolves around the damaging effect that fake news can have on our memories and beliefs, precipitating far-reaching consequences for both individuals and society. Belief in inaccurate information is associated with greater distrust in the media (Tandoc Jr. et al., 2021), reluctance to engage in COVID-19 preventative health behaviors (Freeman et al., 2022), and can even sway critical decisions such as election outcomes (Allcott & Gentzkow, 2017). As a result, the need to find effective strategies to counteract the consequences of fake news exposure is pressing.

I define fake news here as *misinformation* that was mistakenly presented as true by news sources (Lazer et al., 2018; Vraga & Bode, 2020). This is distinct from *disinformation* that is intentionally spread to cause harm (Fallis, 2015). The distinction between these two terms is important as the present integrated dissertation will focus on studies that I conducted to explore the effects of correction methods on misinformation. Cognitive scientists have been actively researching strategies to correct misinformation, but evidence shows that corrections are not always effective in combating the effects of misinformation (for a review, see Lewandowsky et al., 2012). Even after being corrected, misinformation can persistently affect beliefs, reasoning, and actions; a phenomenon known as the continued influence effect (Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988). The effectiveness of corrections can be influenced by various factors (for a review, see Prike & Ecker, 2023), including but not limited to an individual's prior knowledge and beliefs in misinformation (Ecker et al., 2014), credibility of the correction source (Guillory & Geraci, 2013), and level of correction details (for a meta-analysis, see Chan et al., 2017).

One contentious issue in the literature pertains to the content of corrections: specifically, the debate concerns whether it is beneficial or detrimental to include reminders of the misinformation before presenting corrections. This controversy is central here, as it raises critical questions about the effects of misinformation on correction efficacy. The prevailing view was that reminders of misinformation should be avoided (Skurnik et al., 2007; as cited in Schwarz et al., 2007). According to the familiarity-backfire view repeating misinformation during corrections could make it more familiar, potentially increasing its accessibility and perceived accuracy (e.g., Schwarz et al., 2016). However, this view has shifted, and it is now recommended that corrections should include reminders of misinformation (Lewandowsky et al., 2020). Consistent with integration accounts, corrections repeating misinformation can facilitate memory and belief accuracy by increasing the likelihood of noticing conflicts and allowing both representations to be integrated together (for a review, see Ecker et al., 2022).

Here, I will present empirical studies that assessed and compared the predictions stemming from these theoretical frameworks on how reminders, retrieval, and repetitions of fake news influence correction efficacy. The first empirical paper compares the influence of fake news reminders to other correction methods that increase saliency of the fake news details or corrections including real news details on subsequent memory and belief accuracy for news details, underscoring the importance of recollection-based retrieval at test (Kemp et al., 2022a). The second paper examines how detecting corrections and retrieving fake news details during the presentation of corrections influences subsequent memory for real news details, offering theoretical insights into the importance of co-activating fake and real news details during encoding to support later recollection of those details (Kemp et al., 2022b). Finally, the third empirical paper investigates how repeated exposure to fake news before corrections impact memory and belief accuracy in both younger and older adults, offering a more generalizable understanding of how fake news exposure influences a variety of age groups (Kemp et al., 2023). This collection of research expands our comprehension of how retrieval, reminders, and repeated exposure of fake news influences downstream memory and belief accuracy for news headlines. Such findings can provide valuable insights into strategies for mitigating the consequences of fake news in various aspects of everyday life, such as politics and pandemics.

In summary, in the present integrated dissertation, I aim to present research that examines how variations of enhancing accessibility of fake news before corrections influences downstream memory and belief accuracy for news headlines. To do this, I begin with a brief overview on the continued influence effect, discussing how it can be attributed to familiarity misattributions as well as failures to notice conflict and integrate the misinformation and correction. I will then describe the Memory-For-Change framework (Jacoby et al., 2015; Wahlheim & Jacoby, 2013), a perspective that combines mechanisms from both accounts, shedding light on the circumstances in which greater exposure to fake news can either improve or impair memory for corrections. Following this I will discuss the importance of studying how the effects of fake news on memory and belief updating for real news corrections extend to older adults, and the significance of investigating the relationship between memory and beliefs in the context of fake news corrections. After presenting my empirical work, I will conclude with an integrated discussion, exploring the implications for our understanding of the impact of misinformation on memory and belief accuracy, the relationship between memory and beliefs, and charting a path for future research to enhance our scientific understanding of correcting fake news.

Theories Underlying the Continued Influence Effect

The types of correction methods we use to address fake news today are profoundly influenced by our understanding of the continued influence effect; a cognitive phenomenon where people persistently cling to misinformation even after it has been corrected (Lewandowsky et al., 2012). In the initial studies of the continued influence effect (Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988), participants were exposed to a fictional news report describing a warehouse fire that was supposedly caused by volatile materials stored in a nearby closet. Later in the report, this information was debunked and corrected, revealing that the closet was actually empty. To assess the impact of misinformation, participants' understanding of the event was evaluated through a series of free recall and inference questions. The common finding is that despite remembering the correction occurred, participants continue to use misinformation in their inferential reasoning for the cause of the event (for a review, see Lewandowsky et al., 2012).

Recognizing the persistence of false information in individuals' beliefs and memories has driven cognitive scientists to explore factors that influence correction efficacy and the underlying mechanisms. One specific factor believed to impact correction efficacy is the content of the correction, specifically whether it repeats fake news details. It was initially recommended by researchers that corrections should avoid repeating misinformation details (Lewandowsky et al., 2012). The theoretical rationale is that repetition can increase its believability because frequency (Hasher et al., 1977), familiarity (Fazio et al., 2015), and processing fluency (Reber & Schwarz, 1999; Unkelbach, 2007) can be mistakenly used as a signal for truth. Support for this comes from the illusory truth effect showing that statements are perceived as more accurate after repeated exposure (Bacon, 1979; Hasher et al., 1997). The illusory truth effect was initially demonstrated using ambiguous trivia statements (Hasher et al., 1977), but has also been found using statements that are known to be false (Fazio et al., 2015), implausible (Fazio et al., 2019), distributed from a noncredible source (Begg et al., 1992), and with fake news headlines featuring veracity labels (Pennycook et al., 2018).

Expanding on this further is research showing that increasing familiarity with misinformation through its repetition during a correction can sometimes strengthen people's belief in or reliance on the misinformation (Swire et al., 2017a; Weaver et al., 2007). The familiarity backfire effect was first demonstrated by Skurnik et al. (2007; as cited in Schwarz et al., 2007), who exposed participants to a flyer that compared "myths vs. facts" related to the flu vaccine. Upon an immediate test, participants were capable of recalling details that debunked the myths. However, after a 30-minute delay, participants who saw the flyer were more likely to misremember the myths as true and exhibited more negative attitudes towards the flu vaccine compared to participants who had not seen the flyer. To explain this backfire effect, the authors proposed that after a delay participants' memory for the veracity declines, leaving them to rely on the perceived familiarity of the information to determine its perceived accuracy (also see, Begg et al. 1992; Skurnik et al., 2005). This is consistent with the dual-process theory which proposes that retrieval can be based on controlled recollection of contextual information, which can include veracity and source details, or automatic familiarity that reflect memory strength without context (Ayers & Reder, 1998; Jacoby, 1991; Yonelinas, 2002).

Despite some work suggesting that misinformation repetition can lead to a backfire effect (Cook et al., 2014; Pluviano et al., 2019), it is important to note that there is little empirical evidence reporting the effect. Indeed, there have been several failed attempts to observe the effect through replications (Cameron et al., 2013; Wood & Porter, 2019; Ecker et al., 2023), and in situations that are conducive for the effect, such as delay periods between exposure and test (Swire et al., 2017a), and standalone corrections without prior misinformation exposure (Ecker et al., 2020; Prike et al., 2023; but see, Autry & Duarte, 2021). Moreover, recent reviews have suggested that the backfire effect may be attributed to issues with study design and measurement (Swire-Thompson et al., 2020), or due to poor item reliability (Swire-Thompson et al., 2022). Together, the inconsistent evidence suggests that recommendations against misinformation repetition are unwarranted.

In contrast to the familiarity backfire view, other accounts propose that coactivating conflicting information through repeated exposure may enhance memory and belief updating. Ecker et al. (2017) presented participants with fictitious news events containing inaccurate details that were subsequently retracted. Some retractions explicitly repeated the misinformation, included a subtle reminder that the information was incorrect, or did not refer to the misinformation at all before correcting it. When participants were later asked about the events, they were less likely to rely on the misinformation when corrections featured explicit or subtle reminders compared to no reminders, with explicit reminders being the most effective. The authors explained these findings by adopting a conflict saliency account, suggesting that repeating misinformation serves to increase the saliency of conflict. This, in turn, allows individuals to integrate the updated details into their existing mental model of the event. This aligns with research on successful knowledge revision in text refutation studies (Kendeou et al., 2014, 2019).

Compatible with the view that co-activation and controlled retrieval can facilitate updating is the Memory-For-Change account (MFC, e.g., Jacoby et al., 2015; Wahlheim & Jacoby, 2013) that was originally developed to account for how remindings affect episodic memory updating. According to this account, when individuals study new information that shares features with previously studied information, it may trigger a reminder of that prior information (cf. Hintzman, 2011). These reminders aid in detecting changes between events and integrating them into a single memory representation, including their temporal order. Recollecting this integrated representation facilitates memory updating, as individuals can use contextual information to override the automatic influence of the original information (i.e., familiarity). Failure to recollect this integrated representation impairs memory updating, as reminding increases accessibility to the original information through retrieval practice, leading to enhanced reliance on the familiarity of the original information.

In light of this framework, it follows that misinformation reminders could potentially boost memory and belief accuracy for corrections by increasing conflict detection and recollection-based retrieval. This proposition was recently explored by Wahlheim et al. (2020) in their investigation of the effects of reminders on fake news corrections. Their study included multiple phases: In Phase 1, participants assessed the perceived accuracy of real and fake news headlines. In Phase 2, participants saw real news headlines that either reaffirmed the real news headlines or corrected the fake news headlines they saw in Phase 1, with some corrections preceded by reminders of earlier fake news. In the final phase, participants were given a cued recall test for correction headlines from Phase 2, rated them for perceived accuracy, and identified instances when a correction occurred. If a correction was identified, participants were asked to recollect the misinformation. Reminders before corrections significantly improved memory and belief accuracy than corrections without reminders. Moreover, reminders also improved participants' ability to recollect the misinformation, and this was associated with reduced misinformation reliance and enhanced perceived accuracy. These findings add to the growing body of evidence and extend the MFC framework suggesting that misinformation reminders enhance correction efficacy by increasing conflict saliency (Ecker et al., 2017) and memory for misinformation and its subsequent correction.

However, Wahlheim et al.'s (2020) study does have a notable limitation. Reminders of misinformation were consistently followed by real news corrections featuring veracity labels, yet there was no contrast condition in which only real news corrections featured veracity labels. This limitation makes it challenging to disentangle whether the observed effects were primarily driven by improved recollection or whether enhanced conflict saliency played a more prominent role, as suggested by Ecker et al. (2017). The first empirical paper in this integrated dissertation is dedicated to this investigation (Kemp et al., 2022a). In Experiment 1 of this paper, I draw inspiration from the design employed by Wahlheim et al. (2020) but introduce a novel experimental condition in which real news corrections feature veracity labels. This condition allowed us to determine whether the combined presentation of reminders and veracity labels results in memory and belief accuracy improvements beyond what can be achieved by corrections solely incorporating veracity labels. The identification of such additional benefits would imply that reminders exert an influence that transcends the enhancement of conflict saliency—perhaps by encouraging integration—thereby shedding light on the relative roles of conflict awareness and integrative encoding in the effectiveness of corrections.

To foreshadow, I found that reminders that preceded veracity-labeled corrections led to a higher memory and belief accuracy compared to veracity-labeled corrections without reminders.

This suggests that conflict saliency contributes to the benefits of reminders but does not account for all the observed effects. Our conditional results suggest that the added benefits are due to integrative encoding that supported recollection-based retrieval, consistent with the MFC framework. We also found that reminders led to memory impairment when corrections could not be recollected, consistent with the familiarity backfire view. This finding prompted us to consider whether another correction technique that limits the output of fake news is a superior method to providing reminders. Experiment 2 in the first empirical paper of this integrated dissertation aimed to juxtapose the benefits of fake news reminders with another condition, whereby fake news is labeled on its debut, and subjects are instructed to intentionally forget it. This approach was inspired by research showing that memory for recently learned information is better when participants are instructed to forget earlier-learned information (Bjork, 1970; for a review, see Sahakyan et al., 2013). Our results contribute to this evolving literature and may help delineate the most effective strategies for correcting misinformation. If fake news reminders emerge as the superior approach, it underscores their efficacy in promoting accurate memory and belief updates, shedding light on the potential superiority of this correction method.

The findings summarized above suggest that repeated exposure to fake news during corrections may either impair or improve memory for the corrections, depending on whether misinformation familiarity is opposed by recollection-based retrieval (Wahlheim et al., 2020). However, no studies have directly investigated whether self-initiated detection of real news that corrects earlier-studied fake news and retrieval of fake news during the detection process can lead to a mixture of improved and impaired memory for real news details. As mentioned earlier, the MFC framework (Wahlheim & Jacoby, 2013) proposes that when detected corrections are recollected this should be associated with improved memory for real news details by providing more access to integrated representations. Conversely, when detected corrections are not later recollected this should be associated with impaired memory for real news details, as the increased accessibility of the misinformation is unopposed by recollection. The goal of the second empirical paper in this integrated dissertation tested this prediction (Kemp et al., 2022b). That study asked participants to overtly indicate when real news headlines corrected fake news headlines that they studied in an earlier phase and to recall the fake news headlines.

Adult Age Differences in the Effects of Fake News Repetitions

As suggested earlier, research shows that repeated exposure to fake news can impair or improve memory for real news corrections. However, these studies have only focused on younger adults, leaving older adults unstudied. One reason to predict that multiple exposure to fake news will lead to greater impairments for older adults than younger adults is that recollection declines as a function of aging (for a review, see Park & Festini, 2017), whereas familiarity remains relatively intact (Jennings & Jacoby, 1993). This prediction aligns with findings showing that amplifying the familiarity of misinformation through repetition tends to provoke a backfire effect, potentially rendering older adults more susceptible to memory errors due to their diminished reliance on recollection to counter the automatic influences of familiarity (Jacoby, 1999). Support for this assumption comes from work showing that older adults are more likely to make false memory and belief errors under conditions that promote familiarity-based processes. For example, a study examining the mechanisms of belief updating showed that adults ages 65 and older were significantly worse at sustaining their post-correction belief that false claims were inaccurate relative to adults ages 55-64 (Swire et al., 2017a). Likewise, another study revealed that after a delay, older adults were more likely to misremember myths as facts after repeated retractions compared to single retractions (Skurnik et al., 2005).

Another consideration is that older adults might derive benefits from repeated exposure to fake news depending on the extent to which repetitions promote controlled memory processes. The dual-process account of memory holds that older adults experience recollection deficits, diminishing the extent to which current event details cue the retrieval of existing memories and associated contextual information (e.g., Hay & Jacoby, 1999; Jacoby, 1999). This deficiency reduces the opportunities to establish associations across episodes, which hinders the incorporation of critical details, such as the temporal relationship between events (Wahlheim, 2014). To address this recollection deficit, multiple exposures to fake news could play a remedial role by enhancing access to pre-existing memories of the fake news before introducing corrections. This approach may foster enriched and integrative memory representations, facilitating subsequent memory for corrections and its connection with fake news. According to the MFC framework, older adults should benefit when repetitions promote successful detection of corrections and subsequent recollection of misinformation and its correction details. However, despite this potential, older adults might not benefit as much as younger adults due to age-related deficits in associative encoding that undermine their ability to update memories and overcome interference. Evidence to support this is provided by work investigating age differences in memory updating studies using paired associate learning (Wahlheim, 2014) and everyday events (Wahlheim & Zacks, 2019). The goal of the studies in the third empirical paper in this integrated dissertation (Kemp et al., 2023) was to test these theoretical possibilities by examining how repeated exposure to fake news influences memory and belief accuracy for news details in younger and older adults.

Association Between Memory and Belief Accuracy

As mentioned earlier, two of my papers here (Kemp et al., 2022a; Kemp et al., 2023) delve into the impact of fake news on memory and belief accuracy for news details. Examining memory and belief accuracy together is important because there is limited understanding of how these outcomes are related following fake news corrections. Understanding these effects on both measures is important because evidence suggests that beliefs are influenced by memory traces of an event representation (Begg et al., 1992; Berinsky, 2017; Kowalski & Taylor, 2017; Newman et al., 2022; Swire-Thompson et al., 2023; but see, Collier et al., 2023), suggesting that memory may play a central role in belief accuracy. Supporting this idea is evidence from work on the illusory truth effect indicating that recollecting contextual details, such as the statements' prior exposure, can enhance the accuracy of truth judgments (Begg et al., 1992). However, other work indicated that the influence of memory differs depending on whether the task permits participants to adopt an online judgment or memory-based judgment strategy. For instance, a strong correlation was only observed between memory and belief judgments under task conditions where participants were told to make a judgment after stimuli exposure (memory-based) and not when participants were told about the judgment before stimuli exposure (Hastie & Park, 1986). Given that the procedures in the experiments of the present papers require the retrieval of news headline details before making a belief rating about those details, belief ratings are necessarily based on memories.

Within the context of fake news corrections, only a few studies have measured the effects of corrections on memory and beliefs. One study that examined the relationship between memory and belief in the context of correcting everyday misinformation was conducted by Wahlheim et al. (2020). As described earlier, this study examined how reminders of misinformation before the presentation of a veracity-labeled correction influenced subsequent memory for and beliefs in the correction and misinformation intrusions. The results showed an association between memory and beliefs: Participants expressed greater belief that the real news correction details they recalled were true when they could remember that fake news had been corrected. Participants also exhibited a stronger belief that fake news was true when they could not remember that fake news had been corrected. Akin to these findings, Swire-Thompson et al. (2023), showed that the ability to sustain beliefs depended on how well veracity information is represented in memory. Cumulatively, the available literature suggests that memory plays a role in belief accuracy (but see, Collier et al., 2023).

The lack of work examining the relationship between memory and belief accuracy in the context of correcting misinformation underscores the need for attention to this area. I aim to address this gap in the literature through two papers included here. The experiments in the first empirical paper attempt to do this by comparing reminder-based corrections to other correction methods that label the veracity of fake and real news correction on memory and belief accuracy for news details (Kemp et al., 2022a). Additionally, the experiments in the third empirical paper examine how repeated exposure to fake news before corrections influences memory and belief accuracy for both younger and older adults (Kemp et al., 2023). The insights gleaned from these studies may guide current methods on effectively correcting fake news in the real world, fostering sustainable belief change that more accurately informs everyday decisions for individuals of different age groups, such as voting or willingness to vaccinate against COVID-19 and other infectious diseases.

Aims

The primary aim of my research program is to examine how the accessibility of misinformation during the correction process shapes the effectiveness of corrections and the underlying mechanisms at play. Guided by the MFC framework, which posits that memory and belief updating thrive when correction detection and subsequent recollection align but falter when corrections are detected without subsequent recollection, this research endeavors to test the predictions emanating from the MFC framework (Wahlheim & Jacoby, 2013). It aims to investigate how various manipulations altering the accessibility of fake news contribute to differences in correction detection and recollection and how these factors are associated with variations in correction efficacy. Conducting controlled experiments enhances our ability to pinpoint the specific mechanisms that contribute to the effectiveness of corrections. The use of naturalistic stimuli in our research expands the potential for these findings to apply to a broader range of real-world situations. Additionally, my work carries significant implications for social media organizations aiming to combat the effects of misinformation through diverse correction strategies.

Empirical Paper 1, published in *Scientific Reports* (Kemp et al., 2022a), examined the effects of labeling real news corrections to enhance conflict salience (Experiment 1) and labeling fake news on its debut to encourage intentional forgetting (Experiment 2). The findings revealed that reminders yielded the highest memory and belief accuracy, while veracity labels had selective effects. Correction labels resulted in intermediate memory and belief accuracy, whereas fake news labels improved belief accuracy more than memory. Conditional analyses underscored that memory and belief accuracy were superior when participants could recall both real and fake news details, emphasizing the critical role of integrative encoding, most effectively promoted by fake news reminders.

Empirical Paper 2, published in *Cognitive Research: Principles and Implications* (Kemp et al., 2022b), tested predictions from the MFC framework to examine the extent to which detection and recalling fake news during real news corrections impairs or enhances memory updating for such corrections depending on the subsequent use of recollection-based retrieval. Participants overtly indicated when real news headlines corrected fake news headlines studied earlier and recalled the fake news headlines. The relationship between detection, fake news recall, and later memory performance on the test was assessed. Results indicated that detecting corrections and recalling fake news were associated with improved memory for real news when fake news was recalled again and impaired memory when it was not. These findings advocate for a comprehensive theory considering the role of integrated encoding and recollection-based retrieval to account for the contradictory effects of corrections.

Empirical Paper 3, under review at *Cognitive Research: Principles and Implications* (Kemp et al., 2023), examined how repeating fake news before corrections impacted memory and belief accuracy in younger and older adults using recognition (Experiment 1) and cued recall (Experiment 2). Participants were exposed to fake news headlines either three times or once before seeing corrections. Findings demonstrated that neither age nor fake news repetitions affected recognition or cued recall of real news, false alarms to fake news, or perceived accuracy. However, repeated fake news intruded more in cued recall for both age groups. Conditional analyses revealed that fake news repetitions promoted integration for both age groups, enhancing memory accuracy, but also impaired memory accuracy when corrections were not recollected. As expected, older adults remembered fewer corrections and enjoyed fewer associated benefits to real news memory accuracy than younger adults due to age-related recollection deficits. However, they exhibited preserved memory for news headlines via semantic support. These findings underscore that the impact of fake news repetitions on memory depended on age differences in detecting and remembering corrections.

CHAPTER II: FAKE NEWS REMINDERS AND VERACITY LABELS DIFFERENTIALLY BENEFIT MEMORY AND BELIEF ACCURACY FOR NEWS HEADLINES

Abstract

Fake news exposure can negatively affect memory and beliefs, thus sparking debate about whether to repeat misinformation during corrections. The once-prevailing view was that repeating misinformation increases its believability and should thus be avoided. However, misinformation reminders have more recently been shown to enhance memory and belief accuracy. We replicated such reminder benefits in two experiments using news headlines and compared those benefits against the effects of veracity labeling. Specifically, we examined the effects of labeling real news corrections to enhance conflict salience (Experiment 1) and labeling fake news on its debut to encourage intentional forgetting (Experiment 2). Participants first viewed real and fake news headlines with some fake news labeled as false. Participants then saw labeled and unlabeled real news corrections; labeled corrections appeared alone or after fake news reminders. Reminders promoted the best memory and belief accuracy, whereas veracity labels had selective effects. Correction labels led to intermediate memory and belief accuracy, whereas fake news labels improved accuracy for beliefs more than memory. The extent that real and fake news details were recalled together correlated with overall memory and belief differences across conditions, implicating a critical role for integrative encoding that was promoted most by fake news reminders.

Introduction

Fake news refers to stories including verifiably false information presented as true. Although fake news has been around for centuries, it recently gained widespread attention when misinformation about the 2016 and 2020 US Presidential elections, the UK Brexit Referendum, and the coronavirus disease 2019 (COVID-19) spread across social media platforms (Pennycook & Rand, 2021). Fake news exposure can have negative consequences for people and societies, such as when COVID-19 misinformation diminished the willingness to vaccinate and recommend vaccination (Roozenbeek et al., 2020) . These and other threats to public health and democracy emphasize the importance of identifying effective correction methods. Reminding people of real-world fake news before correcting it can substantially enhance memory and belief accuracy (Wahlheim et al., 2020). Additionally, veracity labels about the ground truth of news headlines may reduce false beliefs and sharing behaviors (Morrow et al., 2022). However, we know virtually nothing about how updating of memory and beliefs for factual information compares for correction methods using fake news reminders and veracity labels. The present study addressed this issue by comparing memory and belief accuracy for real news headline details when corrections included fake news reminders, only veracity labels for corrections, or only veracity labels for fake news.

Predictions about memory and belief accuracy under these correction methods can be derived from perspectives on misinformation corrections proposing key roles for familiarity and integration mechanisms. A robust finding that has inspired these existing perspectives originates from studies of the *continued influence effect*. This effect occurs when retracting misinformation does not completely eliminate its influence on event comprehension and reasoning (Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988). This effect may persist when retractions include misinformation, thus increasing misattributions of its familiarity when contextual details are not recollected (Ecker et al., 2010; Lewandowsky et al., 2012). This *familiarity-backfire* view was originally proposed to account for the finding that retractions repeating misinformation increased misinformation-based behavioral intentions after a delay (Schwarz et al., 2007). According to this view, memory and belief accuracy for real news headlines that correct fake news should be better when only veracity labels are provided than when fake news reminders appear before real news corrections because reminders would promote fake news familiarity that could backfire.

Although the backfire view has enjoyed popularity (Cook et al., 2014; Pluviano et al., 2019), many studies have failed to find this effect (Swire-Thompson et al., 2020, 2022). For example, in a study of knowledge revision, beliefs in retracted myths were less sustained relative to affirmed facts after a 3-week delay, but a true backfire effect was not observed because post-retraction beliefs did not regress beyond baseline beliefs (Swire et al., 2017a). Additionally, retractions featuring an explicit misinformation reminder reduced the continued influence effect more than retractions without a reminder(Ecker et al., 2017). According to *conflict salience* accounts of mental-model updating, the misinformation repetition fostered co-activation of the erroneous and correct information, enabling conflict detection and updating of event models and beliefs (Kendeou et al., 2014, 2019). This view is compatible with the assertion that detecting conflict between events is necessary to facilitate memory and belief updating (Putnam et al., 2014; Stadtler et al., 2013; Wahlheim & Jacoby, 2013). Moreover, these findings show how repetition-induced familiarity does not always backfire, thus undermining the prior recommendation to avoid reminders of misinformation (Lewandowsky et al., 2012).

In contrast with predictions from the familiarity backfire view, a recent study showed clear evidence that reminders of fake news can enhance the accuracy of memory for and beliefs in real news corrections (Wahlheim et al., 2020). Participants first read news headlines of unclear veracity then read headlines that affirmed real news and corrected fake news. Some of the corrections were preceded by a fake news reminder, while others were not. Similar to earlier findings (Ecker et al., 2017), reminders improved memory and belief accuracy for real news headlines. These benefits were associated with real news details being recalled more often when fake news details were also recalled. According to the *integrative encoding* view, reminders led both fake and real news detail to be co-activated in working memory. This provided the opportunity for those details to be encoded together into an integrated representation that included information about their veracity and relationship to one another (Kendeou et al., 2014, 2019; Wahlheim et al., 2021). However, a key limitation was that veracity labels appeared with real news corrections that followed fake news reminders, but there was not a contrast condition with only veracity-labeled real news corrections. Thus, the contributions of conflict salience and integrative encoding to reminder-induced benefits could not be separated. If integrative encoding contributes beyond the salience from veracity labels, then memory and belief accuracy should be higher when comparing a fake news reminder condition with a condition including only veracity-labeled corrections without reminders.

An additional objective of the current study was to compare the efficacy of fake news reminders to another veracity-labeling method that has yet to be explored. Studies have explored how correction formats influence memory, showing that ordering of myths and facts has no effect (Swire-Thompson et al., 2021), but labels refuting fake news are more effective when they appear after instead of before or during fake news exposure (Brashier et al., 2021). Related to these findings, memory and belief updating may depend on the extent to which people can disregard veracity-labeled fake news immediately after it appears. This idea is supported by work on directed forgetting showing that under specific circumstances, memory for recently learned information is better when participants are instructed to forget earlier-learned information that can serve as a source of proactive interference (Bäuml et al., 2010; Sahakyan et al., 2013).We addressed this issue here by comparing memory and belief accuracy when fake news is labeled on its debut compared to when it is only labeled when appearing as a reminder. The integrative encoding account predicts that fake news reminders will lead to better memory and belief updating by promoting co-activation, whereas a differentiation view from the context-dependent memory literature (Smith & Vela, 2001) predicts that real news details should suffer less proactive interference when co-activation is prevented. However, labeling fake news on its debut could make it more distinctive and available for integrative encoding.

The benefits of fake news reminders attributed to integrative encoding have been accounted for by a verbal theory proposing that integration enhances recollection-based retrieval of competing details and their relationship (Wahlheim et al., 2021). We evaluated this claim here using a hierarchical Bayesian Multinomial Processing Tree (MPT) approach. MPT modeling can describe the cognitive processes underlying cued recall responses (Jacoby, 1998). We used this approach to estimate the contributions of recollection of headlines' veracity and acontextual familiarity of headline topics to final real news recall. Based on dual-process models of memory (Jacoby, 1991, 1999) and reasoning (Evans & Stanovich, 2013; Kahneman, 2011; Pennycook et al., 2015), we assumed that recalling corrections of fake news required recollection to override the familiarity of fake news.

The Present Study

We conducted two experiments to examine whether the benefits of presenting reminders of fake news immediately before veracity-labeled real news corrections would extend to naturalistic news headline stimuli including both images and text. We also compared the efficacy of reminder-based corrections against veracity-labeled real news corrections without reminders (Experiment 1) and veracity-labeled fake news on its debut (Experiment 2). These comparisons were intended to illuminate the mechanisms underlying fake news reminder effects. Labeling only real news should increase its saliency and signal participants to prioritize remembering it, whereas labeling only fake news could encourage participants to disregard it or make it more distinctive. Regardless of the precise effects of veracity labeling, fake news reminders should better promote integrative encoding by increasing opportunities for co-activation more than veracity labels alone.

We tested this hypothesis using a procedure in which participants first read real and fake news headlines from the internet and indicated their familiarity with and belief in each headline (Phase 1). Participants then read real news headlines that verified real news and corrected fake news from Phase 1 (Phase 2). Finally, participants were given a cued recall test including images from the original headlines. Below the headlines were questions about details that either repeated across phases or were corrected in the second phase. Participants attempted to recall both real and fake news details (when applicable) and indicated their belief in what they recalled as real news (Phase 3). Fake news reminders appeared before some real news headlines labeled as corrections in Phase 2. For other headlines, real news headlines were labeled as corrections in Phase 2 (Experiment 1) and fake news headlines were labeled as misinformation in Phase 1 (Experiment 2). Real news headlines also appeared in Phase 2 as unlabeled corrections of fake news and repetitions of real news from Phase 1. Figure 1 illustrates how headlines appeared in each phase across these within-subjects conditions.

Figure 1. Illustration of the Experimental Designs

Phase 1	Phase 2		Phase 3	Headline Type
	This is misinformation from Phase 1.	This corrects misinformation from Phase 1.		
Many of the recent wildfires in California are caused by bad forest management.	Many of the recent wildfires in California are caused by bad forest management.	Many of the recent wildfires in California are caused by downed electric power lines.	What caused the recent wildfires in California?	Fake News Reminders
In Liberia, less than 38% of young women are able to read at age 18.	- - -	This corrects misinformation from Phase 1. In Liberia, around 60% of young women are able to read at age 18.	In Liberia, what percentage of young women are able to read?	Labeled Corrections (E1 Only)
The United States produces more energy than it consumes, making it energy independent. ⚠ MISINFORMATION	- - -	The United States consumes more energy than it produces, making it energy dependent.	How much energy does the U.S. consume, in relation to how much it produces?	Labeled Fake News (E2 Only)
Hospitals are closing in rural America because they do no have access to high-speed internet.	- - -	Hospitals are closing in rural America because they serve people who are less likely to have health insurance.	Why are hospitals in rural areas closing?	Unlabeled Corrections
The majority of American taxes are spent on social programs such as Medicare.	- - -	The majority of American taxes are spent on social programs such as Medicare.	What type of program does the majority of tax-payers' money go towards?	Repeated Real News

Note. An overview of only the correction headline type conditions Experiments 1 and 2. Phase 1 included real and fake news headlines, most of unclear veracity. Phase 2 included corrections of fake news and exact repetitions of real news headlines with the same picture and wording as in Phase 1. Note that real news headlines are not depicted above to emphasize the distinction among correction methods. The labeled corrections (second row) only appeared in Experiment 1 (E1), and the labeled fake news (third row) only appeared in Experiment 2 (E2). Phase 3 included images that appeared with the headlines from the prior phases and questions about key details that were corrected when headlines appeared as fake news in Phase 1 and corrections in Phase 2. Images were removed from the schematic due to copyright issues.

Based on prior findings showing that labels alone can improve memory and belief accuracy (Brashier et al., 2021; Ecker et al., 2020), we expected that labeling only real news corrections or only fake news would improve memory and belief accuracy by providing details that can be recollected to accept (for real news) and reject (for fake news) headlines. However, presenting reminders before corrections can enhance memory and reasoning beyond labels alone (Ecker et al., 2017). We therefore expected that including fake news reminders before real news corrections would lead to the most accurate memory and beliefs by promoting integrative encoding of representations that best support recollection (Ecker et al., 2017; Kemp et al., 2022b; Kendeou et al., 2014; Sanderson & Ecker, 2020; Wahlheim et al., 2020). To the extent that memory and belief accuracy differ across correction methods, we expected process estimates from the MPT models to show parallel differences in the contributions of recollection. It was unclear whether familiarity would contribute differently across conditions as it could promote correct recall or misattributions of fluently recalled fake news (Skurnik et al., 2007).

Methods

All stimuli, data, and analysis scripts are available here: https://osf.io/zg8yx/. These experiments were approved by the Institutional Review Board at The University of North Carolina at Greensboro (UNCG) and were performed in accordance with relevant guidelines and regulations. Participants were recruited from UNCG, provided informed consent, and received course credit or $10 per hour as compensation.

Participants

The stopping rule for each experiment was to obtain useable data from at least 96 participants. These sample sizes match those from Wahlheim et al. (2020) and were sufficient to detect the smallest effects of interest according to power analyses based on that study for the sample in Experiment 1 (SI4) and on Experiment 1 for the sample in Experiment 2 (SI5). The final sample in each experiment included 96 participants (Experiment 1: 60 women, 34 men, 2 gender diverse) ages 18-33 (M = 19.70, SD = 2.48); Experiment 2: 59 women, 34 men, 3 gender diverse) ages 18-28 (M = 18.95, SD = 1.65). In Experiment 1, data were excluded from 11 participants due to technical issues and one participant who was tested after reaching the target sample (108 participants were tested). In Experiment 2, data were excluded from 18 participants due to technical issues and one participant who was tested after reaching the target sample size (115 participants were tested). We deviated from our pre-registered plan to exclude participants based on failed attention checks and instead controlled for that variable in our analyses (for a detailed rationale, see SI6). Participants were tested individually and received course credit or a $15 gift card as compensation.

Materials and Design

Both experiments included 60 headline pairs from fact-checking websites (i.e., africacheck.org, bettergov.org, politifact.com, snopes.com, statesman.com) each comprising a real and fake news headline on the same unique topic. Fake news headlines included a false detail, and real news headline included a true detail that corrected the false detail. All fake news headlines were originally portrayed by the media as being true. The headline format resembled breaking news updates on internet search engines. Real and fake new headlines about a topic appeared below an image related to the topic.

For counterbalancing, the 60 headline pairs were divided into four sets of 15 and rotated through the within-participant conditions; headline pairs appeared equally often in each condition across participants. Sets included comparable topic variety (i.e., politics, crime statistics, global warming, etc.) and distribution of qualitative and quantitative corrections. Qualitative corrections included changed sentence subjects. For example, the topic about the cause of Californian wildfires included the fake news detail that *bad forest management* was the cause, and the real news detail that *downed electric power lines* was the cause. In contrast, quantitative corrections included changed amounts. For example, topic of the percentage of young women in Liberia who can read at 18 included the fake news detail that it was *less than 38 percent* and the real news detail that it was *around 60 percent*.

Experiment 1 used a within-participants design including a Headline Type variable (Repetition, Unlabeled Correction, Labeled Correction, Fake News Reminder + Labeled Correction). Experiment 2 used the same design, but a Labeled Fake News condition was substituted for the Labeled Correction condition. Each experiment included three phases. Phase 1 included 60 headlines (15 real news; 45 fake news). Phase 2 included 60 real news headlines. Phase 3 included a cued-recall test of the 60 headline topics. Each test item included the picture from the earlier-studied headline above an open-ended question about the detail that was corrected in Phase 2 when fake news had appeared in Phase 1.

Procedure

Stimuli were presented electronically using E-Prime Go software (Psychology Software Tools, 2020). In each phase, stimuli appeared in a fixed random order with the restriction that no more than three headlines from the same condition appeared consecutively. The average list position for each condition was equated to control for serial position effects. Figure 1 illustrates the conditions and procedures described below.

Before Phase 1, participants were told that they would read real and fake news headlines and that they should study them for a later test. Each Phase 1 headline appeared twice each for 8000 ms followed by a 500 ms interstimulus interval (ISI). All 60 headlines appeared once in a first cycle before any headline repeated in a second cycle. On the first cycle, participants indicated their familiarity with each headline story from 1 (Definitely Unfamiliar) to 6 (Definitely Familiar). On the second cycle, they indicated their belief in each headline from 1 (Definitely False) to 6 (Definitely True). Each headline appeared 8000 ms followed by a rating prompt that appeared for 4000 ms. Headlines appeared without labels of their veracity for all items in Experiment 1. However, in the second cycle of Experiment 2, headlines in the Labeled Fake News condition appeared alone for the first 6000 ms and then with a message that the headline was false for the remaining 2000 ms. Participants were told to disregard or intentionally forget these items.

Before Phase 2, participants were told that they would read real news headlines. They were also told that some would repeat real news from Phase 1 and others would correct fake news from Phase 1. They were also told about the experimental conditions and to note when headlines were corrections. Each Phase 2 headline appeared once for 8000 ms (+ 500 ms ISI), including fake news reminders that preceded real news corrections. Headlines in the Repetition, Unlabeled Correction, and Labeled Fake News (Experiment 2 only) conditions appeared without labels of their relationship to headlines in Phase 1. In contrast, headlines in the Labeled Correction (Experiment 1 only) and Fake News Reminder + Labeled Correction conditions appeared with labels indicating whether they corrected or repeated fake news.

Before Phase 3, participants were told that they would recall real news details from Phase 2, indicate if the headlines had corrected fake news, and if so, recall the corrected fake news details from Phase 1 (in that order). They were told that they would also rate their beliefs in the real news details that they recalled from Phase 2. Test cues appeared above a text box until participants typed their recall responses. After attempting to recall the real news detail from Phase 2, participants rated their belief that what they recalled was true in reality from 1 (Definitely False) to 6 (Definitely True) in Phase 3. They then indicated whether the real news in Phase 2 had corrected fake news in Phase 1 by pressing 1 (Yes) or 0 (No). After responding "yes," they were prompted to recall the Phase 1 fake news headline. Note that unlike the previous phases, the cued recall test was self-paced, to avoid placing time pressure on the three possible responses given during each trial.

After Phase 3, participants completed a seven-item cognitive reflection test that consisted of a reworded version of the original three-item task from(Frederick, 2005) and a four-item non-numeric task from(Thomson & Oppenheimer, 2016). Tests scores were the number of questions answered correctly. We report the rationale for including this measure and the results of these exploratory analyses including responses from this measure in Appendix A in the Supplementary Information (henceforth SI) Section 7.3 (i.e., SI 7.3).

Results and Discussion

We performed hypothesis tests using mixed effects models including by-participant and by-item random intercept effects to account for those sources of variability. We describe the statistical methods for all measures in SI 1. We also describe additional exploratory analyses that were not central to the goals of the present study in SI 7. In Phase 1, the baseline measures of familiarity and beliefs indicated that participants perceived real news headlines as more familiar

(SI 2.1) and believable (SI 2.2) than fake news headlines. In Experiment 2, participants believed veracity-labeled fake news headlines far less than all the other unlabeled headlines.

Fake News Reminders Enhanced Real News Recall More Than Labeling Corrections

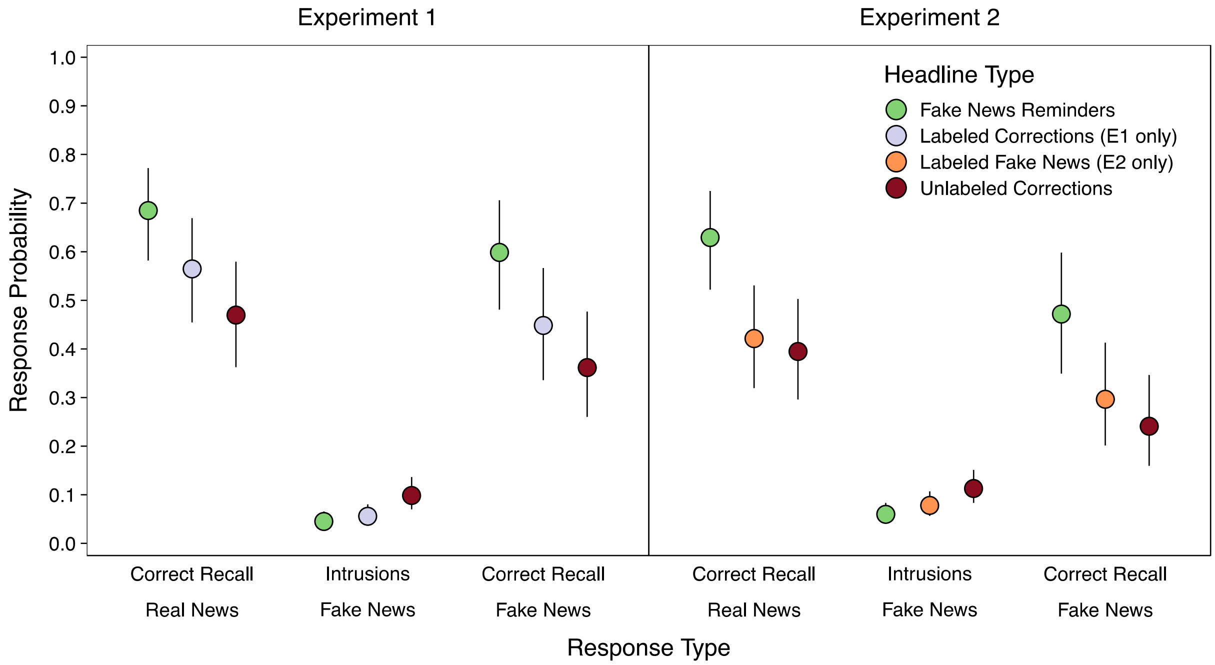
Table 1 displays the complete model results for all analyses of cued recall in Phase 3. Participants recalled real news corrections of fake news in Phase 3 most accurately when fake news reminders had appeared in Phase 2 (Figure 2). Experiment 1 showed significantly higher real news recall when fake news reminders preceded corrections regardless of whether corrections alone were labeled or unlabeled, smallest z ratio = 5.32, $p < .001$. Additionally, real news recall was significantly higher for labeled than unlabeled corrections, z ratio = 4.06, $p < .001$. Experiment 2 showed significantly higher real news recall when fake news reminders immediately preceded corrections than in the other correction conditions, smallest z ratio = 8.91, $p < .001$. Real news recall for unlabeled corrections did not differ based on whether veracity labels accompanied fake news in Phase 1, z ratio = 1.19, $p = .63$. Finally, correct recall for real news that repeated from Phase 1 to Phase 2 (Experiment 1: .76 [*95% CI* = .67, .83]; Experiment 2: .70 [*95% CI* = .60, .78]; not pictured) was significantly higher than for all correction conditions smallest z ratio = 3.28, $p < .01$. Collectively, these results suggest that using fake news reminders to encourage integration of real and fake news promoted real news recall more than increasing conflict saliency for corrections or encouraging participants to disregard fake news with veracity labels.

Table 1. Model Results for Real News Recall, Intrusions of Fake News, and Fake News Recall in Phase 3

		Experiment 1			Experiment 2		
Analysis	Effect	χ^2	*df*	*p*	χ^2	*df*	*p*
Overall Real News Recall	Headline Type	186.74	3	< .001	245.57	3	< .001
Overall Intrusions of Fake News	Headline Type	48.68	2	< .001	34.60	2	< .001
Overall Fake News Recall	Headline Type	97.04	2	< .001	113.31	2	< .001
Conditional Real News Recall	Headline Type	13.40	2	< .01	34.15	2	< .001
	Classification	622.38	2	< .001	635.69	2	< .001
	Headline Type × Classification	3.14	4	= .54	4.06	4	= .41
Conditional Intrusions of Fake News	Headline Type	7.17	2	= .03	4.92	2	= .09
	Classification	84.77	1	< .001	48.66	1	< .001
	Headline Type × Classification	6.64	2	= .04	6.44	2	= .04

Note. The results above correspond to the data visualized in Figure 2 (for overall recall) and Figure 3 (for conditional recall).

Figure 2. Cued Recall of Real and Fake News Details in Phase 3



Note. Probabilities of real news recall, intrusions of fake news, and fake news recall for each correction headline type condition. Points are probabilities estimated from mixed effects models; errors bars are 95% confidence intervals.

Veracity Labels Reduced Intrusions of Fake News

More information about differences in memory accuracy across correction methods can be gleaned from examining intrusions of fake news from Phase 1 during recall of real news from Phase 2 (Figure 2). Memory accuracy on this measure is higher when intrusion rates are lower, indicating fewer memory misattributions. Both experiments showed that labeling corrections of fake news, regardless of whether fake news reminders appeared in Phase 2, led to lower intrusion rates than presenting headlines without labels. Intrusions were significantly higher for unlabeled corrections than for all other corrections, smallest z ratio = 3.55, $p < .01$, and were not significantly different among the other corrections, z ratio = 2.29, $p = .06$. Thus, veracity labels uniformly reduced memory misattributions.

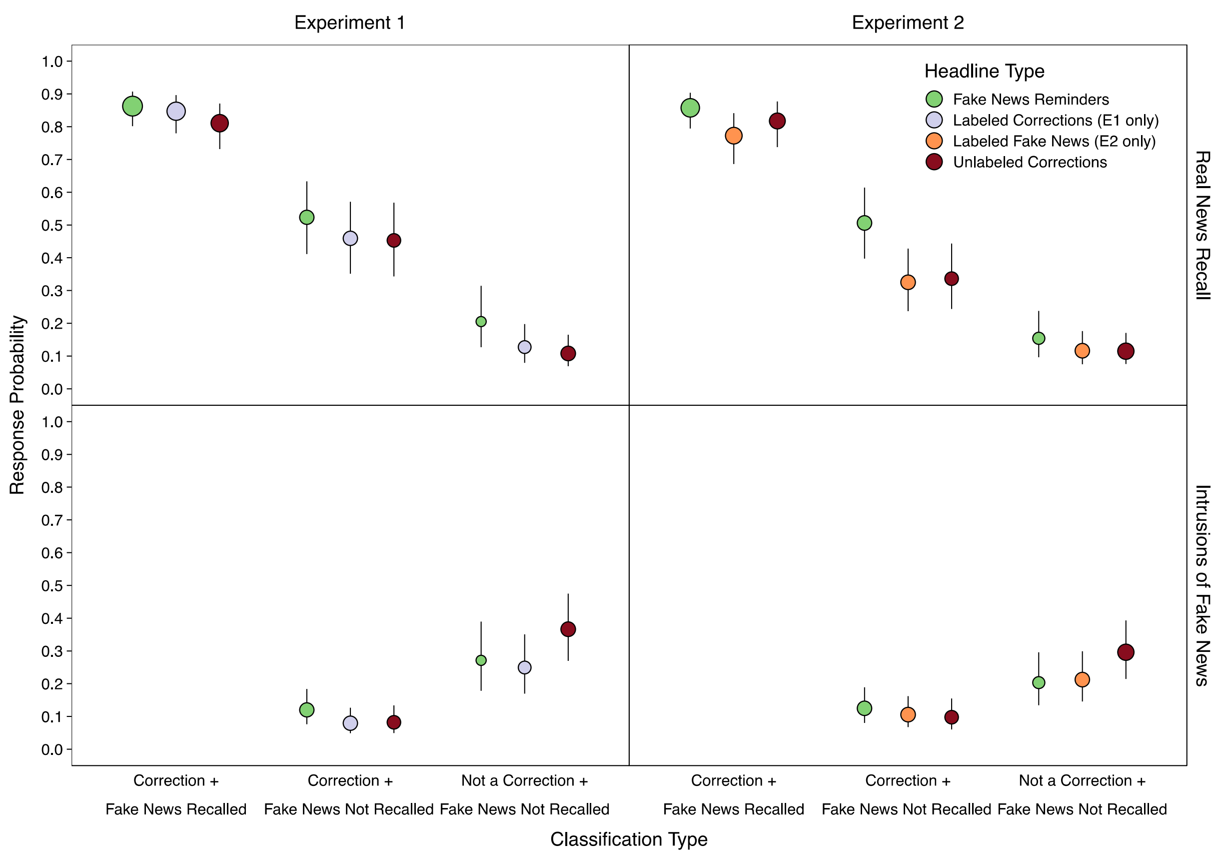
Reminders Enhanced Fake News Recall More Than Veracity Labels

Real news may be better remembered when the details become integrated with the fake news they corrected. We first examined potential associations between fake and real news recall by characterizing the accessibility of fake news across correction conditions (Figure 2). Both experiments showed that providing fake new reminders before labeled corrections led to significantly higher fake news recall than all other corrections, smallest z ratio = 6.24, $p < .001$. Additionally, only labeling corrections (Experiment 1) or fake news (Experiment 2) led to significantly higher fake news recall than presenting corrections without labels, smallest z ratio = 2.88, $p = .01$. These results suggest that repeating fake news as reminders made those headlines most memorable, labeling corrections promoted retrieval practice of fake news when participants thought about what was corrected, and labeling fake news made it more distinctive, despite participants being told to disregard those headlines.

Reminders Promoted Integrative Encoding Over Veracity Labels Alone

We further examined the role of fake news retrieval and integrative encoding during encoding of corrections in memory accuracy for the three correction types in each experiment by computing real news recall conditioned on fake news recall and correction classifications. We created three categories based on combinations of correction classifications and fake news recall (Figure 3). The first two categories included accurately classified corrections that varied based on whether fake news was subsequently recalled. *Correction + Fake News Recalled* refers to headline topics for which participants remembered there was a correction and could recall the fake news detail. *Correction + Fake News Not Recalled* refers to topics for which participants remembered there was a correction and could not recall the fake news detail. *Not a Correction + Fake News Not Recalled* refers to topics for which participants did not remember there was a correction and thus did not recall the fake news detail. Trial proportions corresponding to point sizes in Figure 3 are shown in Table S4.

Figure 3. Conditional Recall of Real News and Intrusions of Fake News



Note. Probabilities of real news recall and intrusions of fake news conditioned on correction classifications for each correction headline type condition. Points are probabilities estimated from mixed effects models; errors bars are 95% confidence intervals. Point sizes indicate for each cell the relative proportion of observations, which are also displayed in Table S4. Values are not displayed for intrusions for classified corrections when fake news was recalled due to sparse observations.

Based on our prior findings (Kemp et al., 2022b; Wahlheim et al., 2019, 2020), we reasoned that integration differences across correction types could be inferred from differences in real news recall probabilities conditioned on fake news also being recalled. In both experiments, real news recall (Figure 3, top panels) was significantly higher for accurately classified corrections accompanied by fake news recall than for other classification types, smallest z ratio = 15.01, $p < .001$, and for accurately classified corrections when fake news was not recalled than corrections that were inaccurately classified, z ratio = 11.21, $p < .001$. Taken with the differences in fake news recall across headline types described above, these findings suggest that real news recall was facilitated to the extent that corrections promoted the co-activation of fake and real news, thus supporting subsequent recollection.

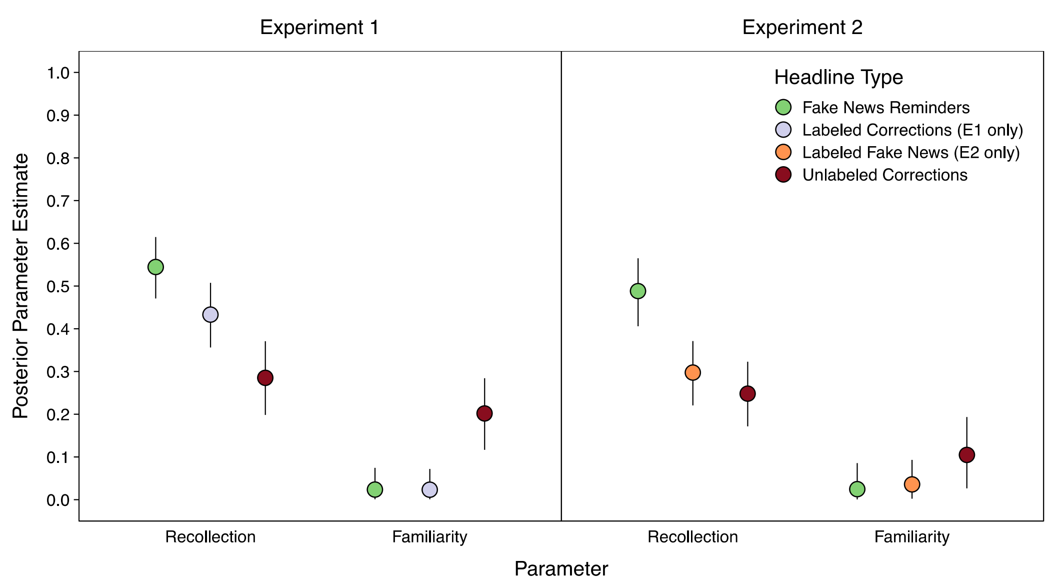
Fake News Intruded more for Inaccurately Classified Corrections

We also examined the extent to which remembering corrections was associated with intrusion reduction, as shown before (Kemp et al., 2022b; Wahlheim et al., 2020). Note that we did not include classifications for which fake news was recalled because intrusions of fake news were redundant responses that almost never occurred. Both experiments showed that intrusions of fake news (Figure 3, bottom panels) were significantly lower for accurately than inaccurately classified corrections. Significant interactions showed that when corrections were inaccurately classified, there were significantly more intrusions for unlabeled than labeled corrections in Experiment 1, z ratio = 2.88, $p = .01$, and unlabeled than both other corrections in Experiment 2, smallest z ratio = 2.58, $p = .03$. These results suggest that remembering that a topic was corrected counteracted familiarity-based misattributions, and this was aided by labels that supported recollection of headline veracity.

Recollection Benefitted more from Fake News Reminders than Veracity Labels

We formally examined the contributions of recollection- and familiarity-based retrieval to cued recall accuracy across correction methods (Figure 4) using the MPT modeling approach explained previously (for a full description of this approach, see SI 3). Recollection estimates when fake news reminders preceded corrections were credibly greater than for all other headline types in both experiments (smallest estimate = 0.11 [0.05, 0.17]). In addition, recollection estimates were credibly greater for labeled than unlabeled corrections (Experiment 1; estimate = 0.15 [0.08, 0.22]), but not credibly different for labeled fake news and unlabeled corrections (Experiment 2; estimate = 0.05 [-0.02, 0.12]). As predicted, these differences paralleled the patterns for correct recall of real news. Familiarity estimates were generally low, but they were credibly greater for unlabeled corrections than all other corrections in Experiment 1 (estimate = 0.18 [0.09, 0.26]), but did not differ across conditions in Experiment 2 (i.e., CIs overlapped with 0). These results support the assertion that the memorial benefits conferred by fake news reminders and veracity-labeled corrections reflect larger contributions of recollection-based retrieval.

Figure 4. Latent Parameter Estimates for Recollection and Familiarity



Note. Parameter estimates for recollection and familiarity for each correction headline type condition. Points are ratings estimated with MPT models, and errors bars are 95% confidence intervals.

Beliefs Distinguished Real from Fake News More with Reminders and Labels

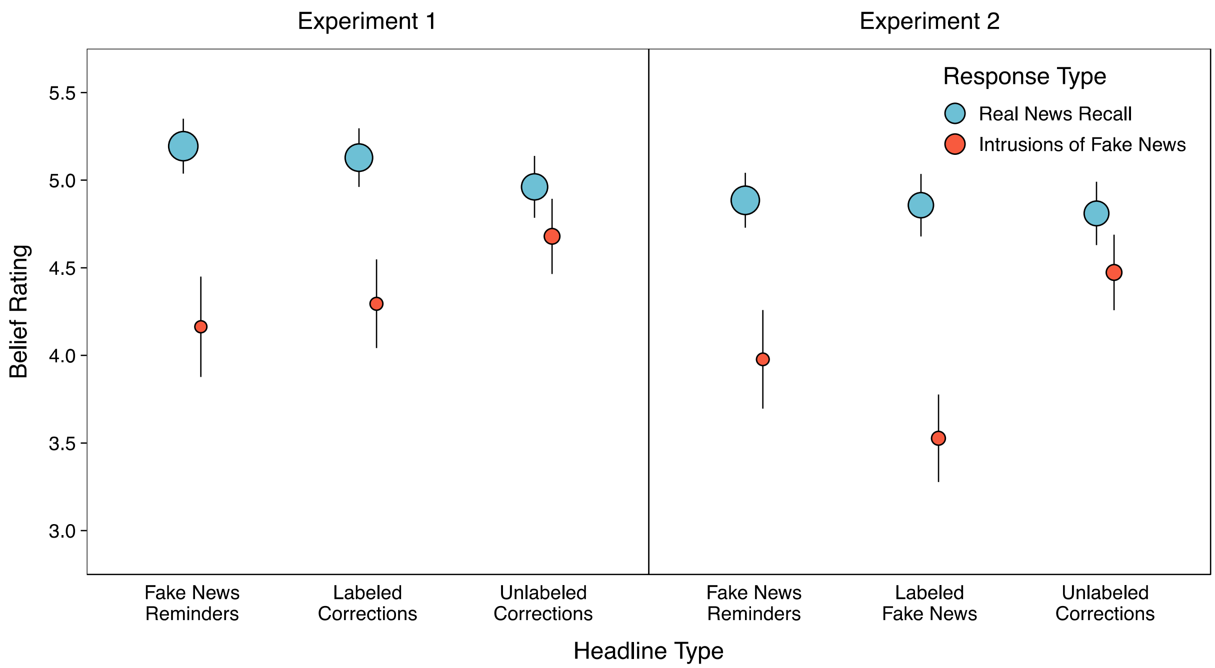
We next examined differences in belief accuracy that were presumably based partly on memory differences across headline types (Figure 5). Table 2 displays the complete model results for all belief rating analyses. We defined belief accuracy as the extent to which ratings were higher for real news recall and lower for intrusions of fake news. We deviated from our preregistered plan by including response type as a predictor instead of assessing each response type separately. Belief ratings were significantly higher for real news recall than intrusions of fake news in both experiments. Significant interactions qualified these differences. Experiment 1 showed significantly higher real news beliefs when fake news reminders had appeared than when corrections were unlabeled, $t(774) = 3.12$, $p < .01$, whereas beliefs in intrusions of fake news were significantly higher when corrections were unlabeled than for other corrections, smallest $t(1322) = 2.81$, $p = .01$. Experiment 2 showed no significant differences in real news beliefs, largest $t(721) = 0.91$, $p = .64$, and significantly lower beliefs in intrusions of fake news for labeled fake news than all other conditions, smallest $t(1233) = 2.68$, $p = .02$, and when fake news reminders had appeared than when corrections were unlabeled, $t(1266) = 3.31$, $p < .01$. These results show that as for cued recall, fake news reminders and veracity labels improved belief accuracy. This conclusion is based on the consistent finding that the difference in belief ratings between real news recall and intrusions of fake news is substantially larger for fake news reminders and veracity-labeled headlines than unlabeled corrections, despite the inconsistency in the pairwise differences for real news recall between experiments. Together, these results suggest that belief accuracy depended partly on recollection of headlines and their veracity.

Table 2. Model Results for Beliefs in Real News Recall and Intrusions of Fake News in Phase 3

		Experiment 1			Experiment 2		
Analysis	Effect	χ^2	*df*	*p*	χ^2	*df*	*p*
Overall	Response Type	67.35	1	< .001	101.84	1	< .001
	Headline Type	1.18	2	= .55	14.29	2	< .001
	Response Type × Headline Type	21.75	2	< .001	32.83	2	< .001
Conditional Real News Recall	Headline Type	4.49	2	= .11	0.25	2	= .88
	Classification	18.06	2	< .001	55.44	2	< .001
	Headline Type × Classification	14.33	4	< .01	4.56	4	= .34
Conditional Intrusions of Fake News	Headline Type	2.61	2	= .27	12.82	2	< .01
	Classification	37.38	1	< .001	26.75	1	< .001
	Headline Type × Classification	3.15	2	= .21	4.43	2	= .11

Note. The results above correspond to the data visualized in Figure 5 (for overall recall) and Figure 6 (for conditional recall).

Figure 5. Belief Ratings of Real News and Intrusions of Fake News



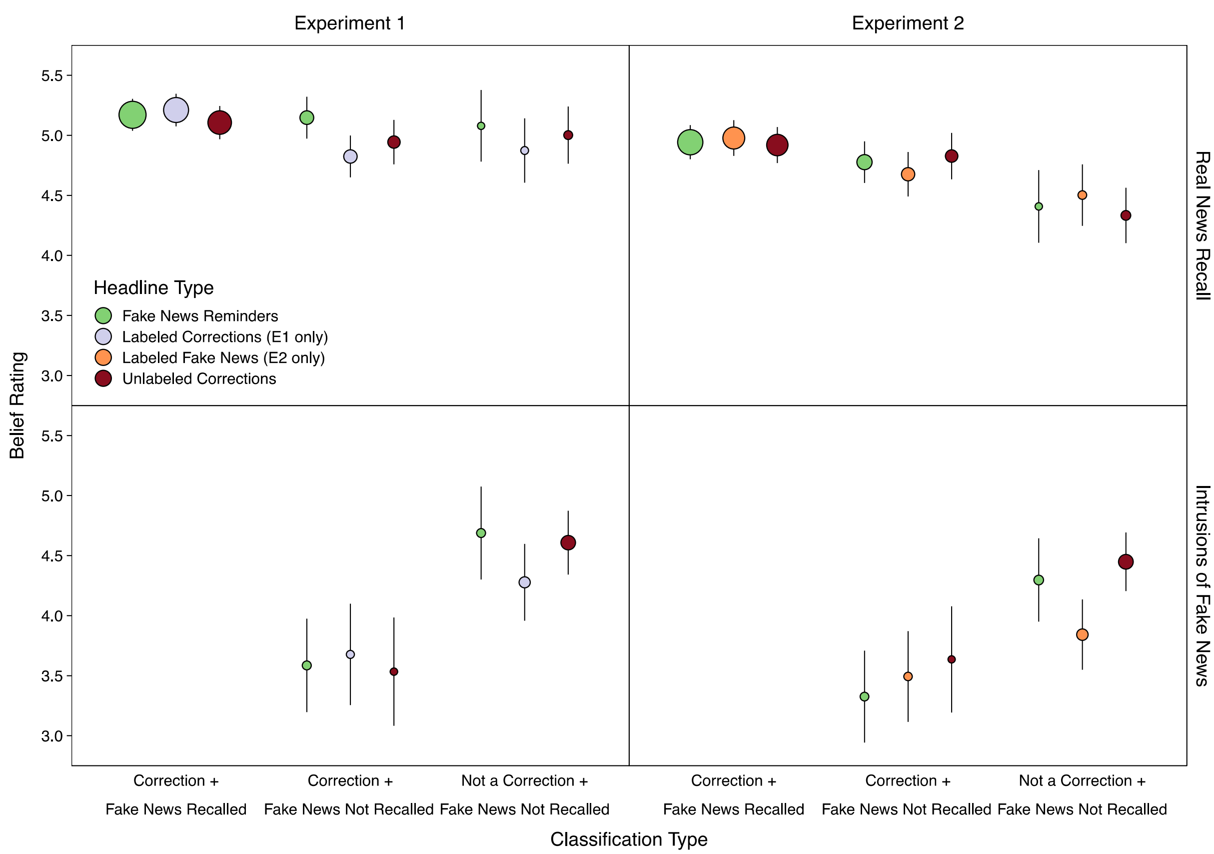
Note. Beliefs in real news recall and intrusions of fake news for each correction headline type condition. Points are ratings estimated with mixed effects models, and errors bars are 95% confidence intervals. Point sizes indicate for each cell the proportion of observations, which are also displayed in Table S5.

Beliefs Better Distinguished Real from Fake News when Corrections were Remembered

We assessed the interplay of memory and beliefs further by conditioning beliefs on correction classifications (Figure 6). Separate models were necessary for each response type because conditional analyses involving intrusions of fake news did not include accurately classified corrections for which fake news was recalled. Table S5 shows the trial proportions. Experiment 1 revealed a significant interaction showing that belief ratings for real news details were consistently high across classifications, except that accurately classified corrections without fake news recall were associated with significantly higher beliefs when fake news reminders had appeared (middle green point) than when corrections were labeled (middle lavender point),

$t(2353) = 3.53$, $p < .01$. Experiment 2 revealed a different pattern. Beliefs in recalled real news were significantly higher for accurately classified corrections with fake news recall than other classifications, smallest *z* ratio = 3.58, $p < .001$, and for accurately classified corrections without fake new recall than inaccurately classified corrections, *z* ratio = 4.09, $p < .001$. Moreover, both experiments showed that belief ratings for intrusions of fake news were significantly lower when corrections were accurately rather than inaccurately classified. These results show that remembering that headline details had been corrected was mostly associated with more accurate beliefs, especially for intrusions of fake news.

Figure 6. Conditional Belief Ratings Real News and Intrusions of Fake News



Note. Beliefs in real news recall and intrusions of fake news for each correction headline type condition. Points are ratings estimated with mixed effects models, and errors bars are 95% confidence intervals. Point sizes indicate for each cell the proportion of observations, which are also displayed in Table S5. Values are not displayed for intrusions for classified corrections when fake news was recalled due to sparse observations.

General Discussion

The present study examined the efficacy of reminder-based and veracity-labeling correction methods for improving memory and belief accuracy for news headlines. These comparisons were intended to identified roles for integration, conflict salience, and intentional forgetting during encoding as well as recollection and familiarity processes during retrieval. Presenting fake news reminders just before labeled corrections improved memory accuracy compared with only labeling corrections or fake news on its debut. Labeling corrections improved memory accuracy compared with presenting unlabeled corrections, whereas labeling fake news conferred no such benefit. Fake news reminders and veracity labels, especially when applied to fake news, both improved belief accuracy relative to unlabeled corrections. Retrieval process estimates and conditional analyses suggested that memory and belief accuracy were better when corrections were recollected. These results suggest that corrections promoting fake news remindings and memory for veracity labels differentially support recollection upon which perceived headline accuracy is based.

The benefits conferred by fake news reminders and veracity-labeled corrections to memory and belief accuracy are compatible with the integration account of the continued influence effect (Ecker et al., 2022). This account proposes that retrieval of outdated information during new learning supports memory updating by promoting conflict saliency and the co-activation of the misinformation and its correction (Ecker et al., 2017; Kendeou et al., 2014; Wahlheim et al., 2020). Support for this account comes from work showing that including misinformation reminders in narrative refutations improves event comprehension (Kendeou et al., 2014) and inferential reasoning (Ecker et al., 2017). Additional support comes from work showing that fake news reminders (Wahlheim et al., 2020) and recall of fake news during corrections (Kemp et al., 2022b) benefit memory and belief accuracy when corrections are recollected. The present findings add to this literature by suggesting that reminder-based and veracity-labeled corrections can promote integrative encoding to the extent that they trigger retrieval of fake news during real news corrections. The present results are also somewhat incompatible with the familiarity backfire prediction that repeating misinformation with corrections should lead misinformation to be more familiar and believable (Schwarz et al., 2016; Skurnik et al., 2007). However, familiarity backfire was likely present in our results when corrections were not recollected. The present findings join the mounting evidence that familiarity backfire in aggregate results is elusive (Lewandowsky et al., 2020; Swire-Thompson et al., 2020, 2022) and provide more evidence for the nuanced interpretation that perceived accuracy is based more on familiarity when testing conditions undermine recollection of corrections (Swire et al., 2017a).

The present results are also relevant for disentangling the mechanisms of fake news reminder benefits. Prior work attributed such benefits to integrative encoding that supported recollection of misinformation, corrections, and their relationship (Wahlheim et al., 2020). However, fake news reminders always preceded veracity-labeled corrections, whereas the contrast condition included only unlabeled corrections. The confound between reminders and veracity labels created ambiguity for interpretation as reminder benefits could have reflected integrative encoding or conflict saliency (Ecker et al., 2017). We eliminated this confound by including veracity-labeled corrections without fake news reminders. Although veracity-labeled corrections improved memory and belief accuracy relative to unlabeled corrections, memory accuracy was greater when fake news reminders appeared. These findings suggest previous reminder benefits reflected contributions of integrative encoding. However, the comparable benefits to belief accuracy of reminders and labels also suggests that recollection of veracity labels are salient cues upon which perceived accuracy is based.

Characterizing veracity label effects on belief accuracy is a focus of the nascent content labeling literature (Morrow et al., 2022). Prior work has shown that veracity labels are more effective at improving belief accuracy when they appear after rather than during or before fake news exposure (Brashier et al., 2021). Our study adds to this literature by showing the consequences for memory and belief accuracy of labeling fake news after exposure and labeling corrections during exposure. Both labels reduced intrusions of fake news and improved belief accuracy compared to when no labels appeared, but real news recall only benefitted when corrections were labeled. These asymmetrical effects suggest that labeling influences recollection of veracity that supports either selecting against false information or selecting for true information. This may explain why labeling fake news mitigated later intrusions but did not enhance recall of corresponding corrections. In this instance, instructions to disregard fake news made those headlines more distinctive, instead of less accessible, contrary to effects sometimes observed in intentional forgetting studies, in which people are instructed to remember some items and forget others (Sahakyan & Foster, 2009). To fully characterize veracity-labelling effects on various aspects of memory and beliefs, future studies should employ other arrangements of labeling, spacing, and repetitions. Studies should also include contextual information in labels, such as news sources and virality measures that provide social feedback (e.g., likes and shares).

Conditional analyses also suggested differences in the extent to which correction methods promoted integrative encoding that supported recollection. Differences in integrative encoding can be inferred from recall of outdated information and the extent to which it is positively associated with memory for updated details (Wahlheim et al., 2021). Here, positive associations between fake and real news recall provided evidence for integration. The memorial benefits associated with fake news recall obtained more often when reminders and corrections were both labeled than when only real or fake news was labeled; these benefits were observed least for unlabeled corrections. This is compatible with the view that conditions that incite looking back to the past enable integrative encoding that supports recollection (Jacoby et al., 2015). Here, reminders appeared to stimulate the most contact between phases, but veracity labels also served this function to a lesser extent. Converging evidence for recollection differences was shown in MPT model estimates as recollection paralleled assumed differences in integrative encoding across conditions.

The finding that recollection estimates were highest in the reminder conditions provides compelling evidence against the familiarity backfire prediction that reinstating fake news should increase the use of familiarity-based heuristics. In fact, familiarity estimates were highest for unlabeled corrections, which were least likely to reinstate fake news during corrections. The present findings align better with the possibility that during encoding, fake news reminders and veracity labels added cues to memory representations that supported recollection rejection (Brainerd et al., 2003; Gallo, 2004; Moore & Lampinen, 2016), which allowed participants to select real news and reject fake news when reporting. As mentioned previously, this may have also improved belief accuracy by allowing cues, such as veracity labels or memory for the relationship between real and fake news, to serve as a basis for judgments. This assertion is supported by the consistently lower beliefs in intrusions of fake news when participants also indicated remembering that fake news had been corrected.

Limitations

As with all studies, the present one had limitations. One aim here was to remove the confounding effect of fake news reminders from the effects of labeling corrections to better account for the role of conflict saliency in correction effects on memory and belief accuracy. However, this does not fully isolate the fake news reminder effect because that would require a condition including fake news reminders alone (i.e., not followed by corrections). In addition, based on visual inspection of the data from both experiments, we decided to include in the analyses participants who failed our benchmark for attention check performance. We mitigated any potential consequences of this by including in each model a by-participant random intercept effect of subjects to account for subsequent memory and belief effects of variability in attention during encoding. Finally, our participants were undergraduates from one university, thus precluding generalizability to the broader population. Future research in this area would benefit from replication attempts using nationally representative samples.

Conclusion

The present study examined the effects of fake news reminders and veracity labels on subsequent memory for and beliefs in real and fake news headline details. Fake news reminders promoted high memory and belief accuracy, consistent with the integrative encoding view and contrary to the familiarity backfire view. Although veracity-labels also enhanced memory accuracy, such improvements were selective and never reached the level promoted by reminders.

However, veracity labelling promoted high belief accuracy suggesting that memory for labels served as a cue for perceived accuracy. Memory and belief differences across corrections largely corresponded with differences in model-derived recollection estimates, which may have characterized the extent to which memory for corrections and associated details were used to select real news and reject fake news. A comprehensive and generalizable understanding of the effects of reminder-based and veracity-labeling correction methods will require examining effects of moderating variables, such as source credibility, headline virality, and political concordance on memory, beliefs, and their relationship

.

CHAPTER III: RECALLING FAKE NEWS DURING REAL NEWS CORRECTIONS CAN IMPAIR OR ENHANCE MEMORY UPDATING: THE ROLE OF RECOLLECTION-BASED RETRIEVAL

Abstract

Fake news can impair memory, leading to societal controversies, such as COVID-19 vaccine efficacy. The pernicious influence of fake news is clear when ineffective corrections leave memories outdated. A key theoretical issue is whether people should recall fake news while reading corrections with contradictory details. The familiarity backfire view proposes that recalling fake news increases its familiarity, leading to interference. However, the integrative encoding view proposes that recalling fake news promotes co-activation and binding of contradictory details, leading to facilitation. Two experiments examined if one theory better accounts for memory updating after participants recalled actual fake news details when reading headlines that corrected misinformation. In Phase 1, participants read real and fake news headlines of unclear veracity taken from various internet sources. In Phase 2, participants read real news headlines that reaffirmed real news and corrected fake news from Phase 1. When they detected that Phase 2 real news corrected fake news, they attempted to recall Phase 1 fake news. In Phase 3, participants first recalled real news details. When they remembered that those details were corrections from Phase 2, they attempted to recall fake news from Phase 1. Recalling fake news when noticing corrections in Phase 2 led to better memory for real news in Phase 3 when fake news was recalled again and worse memory for real news in Phase 3 when fake news was not recalled again. Both views explain part of the memory differences associated with recalling

fake news during corrections, but only when considering whether people recollected that fake news had been corrected.

Significance Statement

The proliferation of fake news in the media can create inaccurate memories that lead to negative effects on public beliefs, decision making, and health. Correcting fake news to mitigate its effects is sometimes effective, but it can also lead to interference in memory when corrections promote the retrieval of misinformation. Corrections facilitate the replacement of outdated misinformation with accurate details (i.e., memory updating), when people can recollect that those discrepancies had occurred. But corrections can also impede such updating when retrieving fake news details during corrections increases the feeling that misinformation details are familiar. The present study shows that these contradictory effects of corrections can be explained by a theory of memory updating proposing key roles for integrative encoding of real and fake news details, retrieval-enhanced familiarity of misinformation, and subsequent recollection of contradictory details and their relationship. The present findings suggest that successfully recalling fake news details when corrections are initially noticed can improve subsequent memory for real news. However, for this to be effective, fake news recalls must enable encoding processes that promote later recollection that fake news had been corrected. These findings support the recommendation that corrections should encourage comparisons of misinformation with accurate information, but only when the relationship between information is encoded well enough to be recollected.

Introduction

The problem of "fake news" is a present dilemma. Fake news is false information presented via media outlets to persuade people that fictional ideas are factual. Exposure to fake news can create memory errors that serve as the basis for inaccurate beliefs. Reported associations between false beliefs and related behaviors suggest that such beliefs could have serious consequences. For example, beliefs based on memory for inaccurate claims about the COVID-19 virus are associated with reduced compliance with health guidelines and reluctance to vaccination (Roozenbeek et al., 2020). To mitigate these consequences, interventions will require clear targets for remediation. Since beliefs are partly based on memory accuracy, interventions will require identifying the mechanisms that allow people to distinguish between episodic memories of false and corrective information.

When corrections appear, the features shared with misinformation often cue retrieval of such information. This leads to repeated exposure that increases misinformation fluency, which can further lead to impaired subsequent memory for corrections (Ecker et al., 2011; Schwarz et al., 2007). These observations have prompted the expert recommendation to avoid reminding people of misinformation during corrections (Lewandowsky et al., 2012) to avoid proactive interference. This view has recently been reversed (Lewandowsky et al., 2020) because research shows that repeating misinformation during corrections can enhance memory and belief accuracy by increasing conflict saliency and enabling integrative encoding (e.g., Ecker et al., 2017; Kendeou et al., 2014). In the present study, we examined how these opposing effects of misinformation retrieval during corrections on subsequent memory accuracy depend on whether participants use recollection-based retrieval. Here, we took a naturalistic approach by examining such mechanisms using stimulus materials comprising actual real and fake news headlines from internet sources.

The idea that retrieving misinformation during corrections can impair subsequent memory for those corrections has been accounted for by dual-process models of memory (e.g., Jacoby, 1991, 1999). Such models propose that retrieval practice increases the accessibility of misinformation by enhancing both recollection of context and acontextual familiarity (e.g., Bishara & Jacoby, 2008). When misinformation is updated by corrections, subsequent memory accuracy for corrections will partly depend on the balance of recollection and familiarity. Recollection of earlier-retrieved misinformation can oppose its familiarity and enable its rejection, whereas the absence of recollection leaves familiarity unopposed and allows misinformation to interfere with correct recall. Thus, retrieving misinformation during corrections should impair memory accuracy, but only when misinformation familiarity is unopposed by recollection-based retrieval (see e.g., Butterfuss & Kendeou, 2020; Ecker et al., 2011). This interplay of retrieval processes is also considered to influence beliefs, such as when repeating information increases perceived truth (i.e., the illusory truth effect). This effect is considered to partly reflect familiarity misattributions (Begg et al., 1992; Schwarz et al., 2007), which is compatible with the view that familiarity can backfire when information sources are not recollected (also see, Skurnik et al., 2007). The illusory truth effect may also arise from repetitions increasing information coherence in semantic memory (Unkelbach & Rom, 2017; Unkelbach et al., 2019). Accordingly, the combination of increased coherence of misinformation from repetitions and decreased coherence of corrections due to its partial repetition of misinformation would lead misinformation to be relatively more fluent. This would impair later recall decisions when retrieved information about corrections is based on coherence and thus perceived truth.

Dual-process perspectives have also been invoked to account for the persistent effects of misinformation on indirect measures of memory and beliefs, such as reasoning, even after a correction (Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988). The *Continued-Influence Effect* (CIE) has been shown in paradigms wherein participants read a fictitious news report that contained a critical piece of information that was subsequently corrected or not corrected. The CIE is observed when corrections reduce, but do not eliminate misinformation reliance. One suggestion for why corrections are not completely effective is that they sometimes repeat misinformation, thereby increasing the potential for familiarity to exert an unwanted influence. However, the view that familiarity will backfire has recently been contested because the available literature shows that such effects are rare and depend on experimental design and stimulus characteristics (Swire-Thompson et al., 2020, 2022).

Evidence for the role of familiarity in the CIE has been recently shown in a study of the mechanisms of belief updating (Swire et al., 2017a). Participants were presented with a series of true or false claims of unclear veracity and rated their beliefs in each. In the next phase, true claims were affirmed, and false claims were corrected. In a final phase, participants re-rated their belief in each statement either immediately or after a delay. Corrections reduced myth beliefs, but such reductions were smaller at longer delays. These findings were interpreted as showing that corrections of myth beliefs were sustained less effectively at longer delays because recollection was less available to oppose familiarity-based judgments. Although this implicates a role for familiarity in myth beliefs, this belief regression was not a true familiarity backfire effect because post-correction beliefs remained lower than pre-correction beliefs. Thus, familiarity backfire did not fully account for the CIE.

In contrast to the view that repeated exposure to misinformation impairs memory for and beliefs in corrections, others have suggested that it enhances updating by promoting awareness of information conflict, which provides the opportunity for integrative encoding. For example, in the traditional CIE paradigm where participants read narratives about events including incorrect details that are corrected in a subsequent narrative before an inferential reasoning test, corrections featuring an explicit misinformation reminder reduced the CIE better than corrections without a reminder (Ecker et al., 2017). The authors proposed that reminder benefits occurred because repeating misinformation fostered its co-activation with its correction. This presumably increased the salience of conflict between the competing information and supported integrative encoding and knowledge revision (Kendeou et al., 2014, 2019). These findings are consistent with work showing that detecting conflict can facilitate memory and belief updating (Putnam et al., 2014; Stadtler et al., 2013). These findings are also compatible with a retrieval account positing that conflict salience enhances the encoding of corrections, thus supporting their later recollection (Ecker et al., 2010; Seifert, 2002).

The findings above suggest that a comprehensive explanation of when misinformation retrieval during corrections will impair or improve memory requires considering the mechanisms proposed by familiarity backfire and integrative encoding accounts. The Memory-for-Change framework (MFC; Jacoby et al., 2015; Wahlheim & Jacoby, 2013) encompasses ideas from both accounts. The MFC proposes that stimulus features from a current event can cue retrieval of an earlier event, which enables detection of changed features between events. The co-activation of events then enables integrative encoding that supports subsequent source memory for event order when retrieval is recollection-based. However, the account also proposes that prior-event retrieval cued by current event features increases the familiarity of earlier events, which can lead to proactive inference effects when event changes are not subsequently recollected—a type of familiarity backfire. This mixture of effects has been shown consistently across paradigms using stimuli varying in naturalism (for a review, see Wahlheim et al., 2021). The generalizability of these findings suggests that whether retrieval of misinformation during corrections impairs or improves subsequent memory for corrections may similarly depend on whether recollection is used at retrieval.

The role of recollection-based retrieval in the effects of misinformation repetition on memory for corrections was recently shown using real-world news headlines (Wahlheim et al., 2020). Participants read real and fake news headlines of unclear veracity, and then read headlines that affirmed the factual headline or corrected the misinformation. Some of the corrections were immediately preceded by a fake news reminder, while others were not. Fake news reminders enhanced overall recall of details from corrective headlines, which supported an integrative encoding account and contradicted a familiarity backfire view. Importantly, conditional analyses of recall showed that these memory benefits reflected greater recollection-based retrieval. In contrast, familiarity backfired when fake news was not recollected in the form of lower correct recall of real news headlines and more false recall of fake news headlines. These findings suggested that fake news reminders can counteract the persistent effect of misinformation by cuing retrieval of earlier fake news and promoting subsequent recollection of the conflict between fake and real news. However, the provision of fake news reminders precluded direct examination of the role of retrieving fake news details when detecting real news corrections. We addressed this limitation for theoretical interpretation here asking participants to overtly indicate when real news headlines corrected fake news headlines that they studied in an earlier phase and to recall the fake news headlines. Using this approach, this study is the first to our knowledge to directly examine associations between fake news retrieval when initial detecting real news corrections and subsequent memory for news details and their veracity.

Since the accessibility of earlier studied misinformation should determine how often corrections are detected as such, manipulations of misinformation memorability should lead to differences in subsequent memory effects associated with retrieval practice of fake news when detecting corrections. One way to influence misinformation accessibility is to vary the congruence of participant and peer beliefs in the veracity of such information (Schwarz et al., 2016). From one perspective, when evaluating the veracity of information, people incorporate peer beliefs into their evaluations, especially when the information has ambiguous veracity (Gabbert et al., 2007) and social contacts endorse the belief (Galesic et al., 2021). In addition, when new information matches prior beliefs, encoding is more fluent (Schwarz et al., 2016), leading to stronger memory representations (Levine & Murphy, 1943), possibly by integrating information with schemas (Pratkanis, 1989). Accordingly, misinformation that both participants and peers believe would be more accessible, leading to better detection of contradictory details enabled by misinformation retrieval. This accessibility would also increase the risk that familiarity would backfire later when recollection is not engaged.

However, belief congruence does not always enhance memory, such as when contrasting it with belief-incongruent information leads the incongruent information to garner more attention (e.g., Maier & Richter, 2013). Related work has shown that incongruence between participant and peer beliefs may also increase misinformation accessibility when contradictions prompt critical processing of that information (Munnich et al., 2007) possibly in reaction to prediction errors piquing interest and attention (Vlasceanu et al., 2021). This would lead to stronger misinformation representations that better support memory for contradictory facts but also pose a greater risk for familiarity backfire when recollection fails. In our second experiment, we conducted an exploratory investigation on how the congruence between participant and peer beliefs affects misinformation encoding and subsequent memory for contradictory facts.

The Present Study

The findings summarized above suggest that repeated exposure to fake news during corrections could impair or improve memory for corrections, depending on whether misinformation familiarity is opposed by recollection-based retrieval. However, no studies have directly examined how detection of real news that corrects earlier-studied fake news and retrieval of fake news during such detection leads to a mixture of improved and impaired memory for real news details. As mentioned earlier, our approach is inspired by work showing that memory, reasoning, and beliefs were improved when explicit reminders of misinformation were provided with corrections (Ecker et al., 2017; Wahlheim et al., 2020). However, those experiments did not directly assess the role of misinformation retrieval cued by corrections with shared features. The present study contributes to theory by illuminating how verified retrievals of fake news when detecting corrections is associated with subsequent memory updating. On the practical side, this experiment is more likely to resemble situations in everyday life in which people encounter corrections without the original fake headline or fact-checker tags.

This gap in the literature was addressed using a variant of the misinformation correction paradigm from Wahlheim et al. (2020). Participants first studied real and fake news headlines of unclear veracity. Next, they studied real news headlines that either reaffirmed the real news or corrected fake news. While studying these headlines, a prompt appeared asking participants to indicate instances when real news headlines corrected fake news and attempted to recall the fake news details when indicated as such. In the final phase, they attempted to recall real news details, indicated if those details had earlier corrected fake news, and if so, attempted to recall the fake news details.

Based on previous findings from similar paradigms in the episodic memory updating literature (for a review, see Wahlheim et al., 2021), we predicted that recalling fake news details when detecting real news corrections would facilitate memory for real news details when the fake news details could be subsequently recollected as having been corrected. In contrast, we predicted that recalling fake news details when detecting real news corrections would interfere with subsequent memory accuracy when the fake news details could not be subsequently recollected as having been corrected. This would occur because the familiarity and thus perceived truthfulness of misinformation would be unopposed (i.e., backfire), leading to poorer recall of correct details and more misinformation intrusions. These findings would be consistent with the MFC framework which subsumes familiarity-based and integrative encoding accounts of the continued influence of misinformation and posits a key moderating role for recollection-based retrieval. We also explored how congruence between peer and participant beliefs during initial fake news encoding interacted with this mixture of effects. We describe potential outcomes for this exploratory aim below when introducing the second experiment.

Experiment 1

Experiment 1 is the first to our knowledge to characterize proactive effects of fake news exposure on memory for real news details when participants initially retrieved fake news details while studying corrections and subsequently recollected such details on a final memory test. This allowed us to determine when prompting participants to retrieve earlier-studied fake news details leads to the positive effects of memory updating as well as the negative effects of familiarity backfire. Importantly, recent research suggests that both mechanisms can contribute to overall memory performance, depending on whether retrieval of real news details is recollection-based (Wahlheim et al., 2020). Based on that research, we expected that recalling fake news details during corrections would lead to proactive facilitation in memory for real news when fake news is subsequently recollected as being corrected and proactive interference when fake news is not recollected. The balance of these effects, along with instances where fake news is not recalled, should therefore determine the direction and magnitude of proactive effects of memory for real news corrections in unconditioned summary scores.

Method

The stimuli, data, and analysis scripts are available on the Open Science Framework (OSF) website: https://osf.io/xnvrj/. The research reported here was approved by the Institutional Review Board at The University of North Carolina at Greensboro (UNCG).

Participants

The participants were 48 UNCG undergraduates (25 women, 23 men), ages 18-25 (M = 19.23, SD = 1.32). Our stopping rule was to test as many participants as possible in approximately one semester given lab resources with the final sample being a multiple of the three experimental formats. We were primarily interested in how proactive effects of memory differed based on whether fake new details that were recalled during corrections were subsequently recalled in Phase 3. However, we also conducted a sensitivity analysis after collecting the data to estimate the statistical power to detect the smallest effect size (odds ratio[*OR*]) of interest in Experiment 1. This corresponded to the difference in real news recall in the Control and Correction headline conditions (*OR* = 1.36). The analysis indicated that

Experiment 1 had 71% power to detect an odds ratio of 1.36. A complete description of this analysis and its results can be found in Appendix B in SI 8.1.

Design

The experiment used a 3 (Headline Type: Repetition vs. Control vs. Correction) within-participants design.

Materials

We selected from various internet sources 45 headline pairs that each included one fake news headline and its real news correction. The headlines were taken from sources such as news center websites (i.e., MSNBC, CBS News), well-known people (President Donald Trump, Bernie Sanders), government statistics websites, and folk myth websites. The fake news headlines were factual errors, and the real news corrections included factual details that contradicted the errors. All fake news headlines were originally portrayed as being true. Each headline pair corresponded to a unique topic. For example, the topic on the US president who took the most vacation days included the fake news headline, “President Obama took fewer vacation days than any other recent president,” and the real news headline, “President Clinton took fewer vacation days than any other recent president.” When preparing the experiment, we were able to find 41 topics related to US events described in various news sources. To increase the number of topics, we also included 4 urban myths (e.g., “Only older people need a flu vaccine.”). To foreshadow, when preparing Experiment 2, we found enough US events to replace the urban myths used here. Both experiments showed the same patterns related to the main hypotheses of this study, thus mitigating concerns about headline-specific effects.

For counterbalancing, we randomly divided the 45 headline pairs into three groups of 15. We rotated groups through Headline Type conditions, which created three experimental formats.

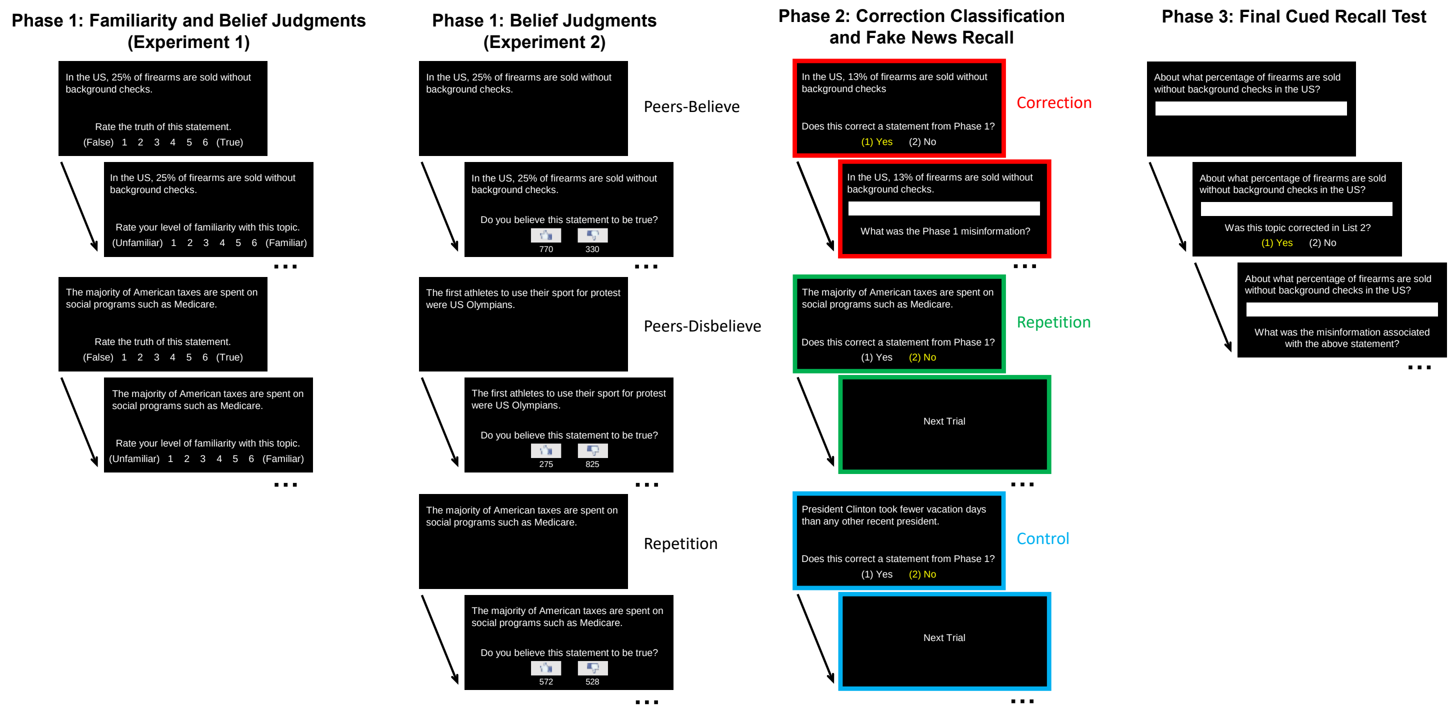
Groups appeared equally often in each condition across participants. Note that Phase 1 included a mixture of fake and real news headlines depending on the condition, whereas Phase 2 included all real news headlines. Phase 1 included 30 headlines (15 fake news; 15 real news). Phase 2 included 45 headlines [15 real news from Phase 1 (Repetition); 15 real news that corrected fake news from Phase 1 (Correction); and 15 real news that appeared only in Phase 2 (Control)]. We included control headlines as a contrast condition against which to assess proactive effects of fake news exposure on memory for real news headlines.

Phase 3 included 45 questions corresponding to Phase 2 headlines that could be answered with either fake or real news details. For example, the question, "Which recent president took the fewest vacation days?" could be completed with the fake news detail "*President Obama*" or the real news detail "*President Clinton*."

Procedure

Participants completed the experiment individually in a quiet room with an experimenter present. Stimuli were presented electronically using E-Prime 2.0 software (Psychology Software Tools, 2012). Participants completed three phases (Phase 1, Phase 2, and Phase 3) in one session. Figure 7 displays a schematic of the Experiment 1 procedure.

Figure 7. Schematic of the Procedure: Experiments 1 and 2



Note. A schematic overview of the trial structures from the procedures in both experiments. The main difference between experiments was the trial structure in Phase 1: In Experiment 1, participants rated their familiarity and believability of headlines; In Experiment 2, participants rated their believability of each headline, which displayed the number of fictional peers who believed and disbelieved the headline. The majority of peers believed the headline in the *Peers-Believe* condition, and the minority of peers believed the headline in the *Peers-Disbelieve* condition. Phase 2 included *Correction* headlines that corrected fake news from Phase 1 (red borders), *Repetition* headlines that repeated real news from Phase 1 (green borders), and *Control* headlines that only appeared

Phase 2 (blue borders). Note that all the trials in the *Peers-Believe* and *Peers-Disbelieve* conditions were later corrected, whereas that Phase 1 trials in the *Repetition* condition always included a negligible difference in peer beliefs. In both experiments, during Phase 2, participants indicated when they detected headlines that contradicted fake news, and if so, attempted to recall fake news from Phase 1. In the first slide Phase 2 trials, the yellow highlights for the “Yes” and “No” judgments indicate the correct classification of each headline type upon which the second slide was contingent. During Phase 3, participants first recalled Phase 2 real news details, then indicated those for which fake news was corrected in Phase 2, and for those, attempted to recall the Phase 1 fake news.

Prior to Phase 1, participants were told that their tasks would be to study individual headlines for an upcoming memory test and to rate the truthfulness and familiarity of each headline. In Phase 1, headlines appeared in a random order for 8000 ms each in white font against a black background. Two prompts then appeared below each headline sequentially for 5000 ms each. The first prompt asked participants to rate the truthfulness of the headline on a scale from 1 (False) to 6 (True). The second prompt asked participants to rate their familiarity with the headline on a scale from 1 (Unfamiliar) to 6 (Familiar). Participants were encouraged to use the full range of the scale and made their ratings by pressing a number on the keyboard. The prompt color changed from white to yellow after participants entered their ratings. Each slide was followed by a blank screen for 250 ms. The purpose of including these judgments was to obtain baseline assessments of beliefs and familiarity for all items and to give participants a task that would keep them actively engaged.

Prior to Phase 2, participants were told that some of the headlines they studied in Phase 1 were fake news and that corrections to those headlines would appear in Phase 2. They were also told that Phase 2 would include repetitions of real news headlines from Phase 1 and real news headlines that only appeared in Phase 2. Participants were told to study the headlines carefully for a later test and to indicate which headlines were corrections of Phase 1 fake news. During Phase 2, headlines appeared individually in a random order for 8000 ms each. Next, a prompt appeared below each headline asking participants to indicate if the headline corrected a headline that they studied in Phase 1. Participants made self-paced judgments, responding either Yes (1) or No (2) by pressing a number key. When they responded "Yes," another prompt appeared above a text box wherein they typed only the fake news details from Phase 1 that were corrected in Phase 2 and pressed "Enter" to advance to the next headline. When they indicated "No," the program advanced to the next headline. Note that the participants' judgments were self-paced and that the headlines stayed on the screen while participants made their responses. A mixed-effect model with Participants and Items as random intercept effects indicated that the estimated marginal mean of yes/no durations was significantly lower for Control (*M* = 1587 ms, *95% CI* = [1297, 1877 ms]) than Correction (*M* = 1969 ms, *95% CI* = [1679, 2259 ms]), $t(2066) = 2.91$, $p = .01$, but not Repetition (*M* = 1853, *95% CI* = [1564, 2143 ms]) headlines, $t(2066) = 2.03$, $p = .11$. Durations for Correction and Repetition headlines were not significantly different, $t(2066) = 0.88$, $p = .65$.

Prior to Phase 3, participants were told that their task would be to answer questions about the real news headlines they had just studied in Phase 2. During Phase 3, questions appeared individually in a random order above a text box until participants entered their response. After participants attempted to recall the real news details from Phase 2, the screen cleared, and a prompt asked whether the detail they typed had earlier corrected fake news from Phase 1. They responded by either Yes (1) or No (2) by pressing a number key. When they indicated "Yes," a prompt asked participants to type in the fake news details from Phase 1 that were corrected in Phase 2. When they indicated "No," the program advanced to the next item. Participants were encouraged to respond accurately to all items and were allowed to pass when they could not think of a response.

Statistical Methods

In both experiments, we performed all statistical tests using R software (R Core Team, 2021). To examine the effects of varying headline types, we fitted linear and logistic mixed-effects models using the *glmer* function from the *lme4* package (Bates et al., 2015). Based on Signal Detection Theory (SDT; Green & Swets, 1966) we also characterized participants' detection of corrections during Phase 2 and subsequent memory that corrections had been detected during Phase 3 in terms of discrimination (*d'*) and response bias (*c*). We calculated these parameter estimates by computing hit and false alarm rates for each participant and using the *dprime* function from the psycho package (Makowski, 2021) to estimate the parameters. We performed Wald's χ^2 hypothesis tests using the *Anova* function of the *car* package (Fox & Weisberg, 2019) and posthoc comparisons using the Tukey method in the *emmeans* package (Lenth, 2021).

All models included Headline Type as a fixed effect as well as Participants and Items as random intercept effects to increase power (Miller et al., 2020). Given the self-paced feature of misinformation recall during Phase 2 study, we attempted to control for encoding time differences in the mixed-effects models of Phase 3 recall performance that could have been affected by encoding time differences by including encoding time as a random effect. However, 11 out of 12 models would not converge. For the one model that did converge, the pattern of results was the same as when encoding time was not included. Consequently, encoding time is not included in the models reported below. The complete model specifications are available in the analysis scripts on the OSF: (https://osf.io/xnvrj/). The significant level was $\alpha = .05$.

Results

Familiarity and Belief Ratings (Phase 1)

Participants' familiarity with and beliefs in Phase 1 headlines were compared for real and fake news headlines by fitting separate models to each measure with Headline Type as a factor. The model for familiarity ratings indicated that participants were comparably familiar with real news (M = 2.38, *95% CI* = [2.06, 2.69]) and fake news (M = 2.45, *95% CI* = [2.14, 2.76]), $t(1238) = 1.04$, $p = .30$. In contrast, the model for belief ratings indicated that participants believed real news (*M* = 3.37, *95% CI* = 3.18, 3.57] more than fake news (*M* = 3.08, *95% CI* = [2.88, 3.27]), $t(1241) = 3.74$, $p < .001$. Note that the familiarity ratings were modest relative to the maximum possible rating, indicating that participants knew little about many of the headlines before entering the experiment.

Cued Recall Performance (Phase 3)

We examined proactive effects of fake news exposure on memory for real news by examining correct cued recall of headline details from Phase 2. Cued recall responses were coded independently by two raters who were blind to the experimental conditions. Responses were considered correct when they included real news details from Phase 2 and intrusion errors when they included fake news details from Phase 1. These two response types are key measures of proactive effects of memory. Facilitation effects can be assessed in correct recall of recent information and interference effects can be assessed in both correct recall of recent information and errors originating from a non-recent source. Participants made other types of errors that were not of theoretical interest, so we do not report them here (see SI 9 for a description of all response types and the scoring method used to classify responses).

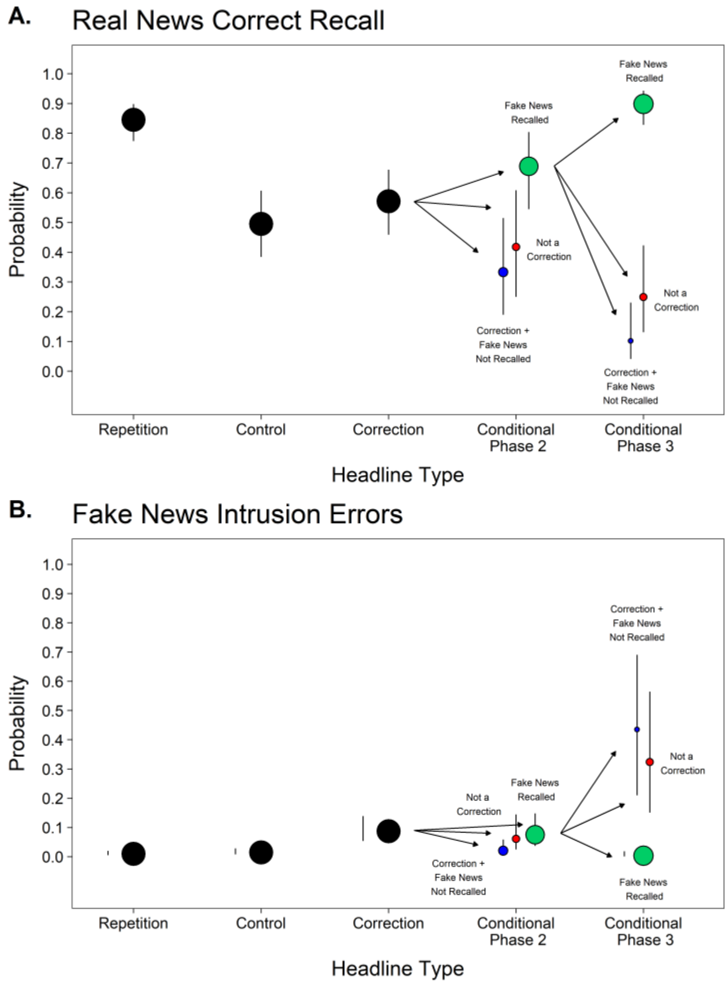
Real News Correct Recall

Figure 8A (black points) displays the overall probabilities of real news correct recall in Phase 3. A model including Headline Type as a factor indicated a significant effect, $\chi^2(2) = 159.32$, $p < .001$. Correct recall was significantly greater for Repetition than Control and Correction headlines, smallest *z* ratio = 10.09, $p < .001$, and for Correction than Control headlines, *z* ratio = 2.48, $p = .04$, indicating overall proactive facilitation in memory for real news that corrected fake news, and thus, no overall familiarity backfire effect.

Fake News Intrusion Errors

Figure 8B (black points) displays the overall probabilities of fake news intrusion errors in Phase 3. Note that intrusions for headlines in the Correction condition reflect episodic memory errors that occur when participants mistakenly report fake news details when asked to recall real news details. In contrast, intrusions for headlines in the Repetition and Control conditions are semantic memory errors that occur when participants spontaneously output fake news details that had not appeared in the experimental context. Therefore, the extent that there are more intrusion errors for Correction than Repetition and Control conditions indicates the contribution of proactive interference from fake news headlines that appeared in Phase 1. A model including Headline Type as a factor indicated a significant effect, $\chi^2(2) = 96.54$, $p < .001$. Intrusions were significantly higher for Correction headlines than Repetition and Control headlines, smallest z ratio = 7.50, $p < .001$, and not significantly different for Repetition and Control headlines, z ratio = 1.24, $p = .43$. These findings show that fake news exposure in Phase 1 led to proactive interference when participants attempted to recall of real news details from Phase 2.

Figure 8. Recall of Real News and Intrusions of Fake News: Experiment 1



Note. Probabilities of real news correct recall (Panel A) and fake news intrusion errors (Panel B) as a function of Headline Type in Experiment 1. Black points represent probabilities for all observations. Colored points represent probabilities conditioned on correction classification types in Phase 2 and 3. The cells represent corrections that were classified as such and for which fake news was recalled (green points), corrections that were classified as such and for which fake news was not recalled (blue points), and corrections that were not classified as such (red points). The size of each point indicates the relative proportion of observations in each cell. Error bars are 95% confidence intervals and are displayed adjacent to the points when the intervals lengths are shorter than point diameters

Correction Classifications and Fake News Recall (Phases 2 and 3)

To identify the role of detection of fake news corrections and recollecting that fake news had been corrected in cued recall accuracy in Phase 3, we first assessed participants' awareness of and memory for such corrections in Phase 2 and Phase 3, respectively. We first computed the separate probabilities of participants indicating during Phase 2 that "yes" a headline was a correction and that during Phase 3 that "yes" the topic had earlier included a correction for all headline types. This provided a general indication of how well participants could discriminate Correction headlines from other headline types. Note that the extent to which participants incorrectly indicated that topics from the Repetition and Control conditions were associated with fake news corrections can be considered one index of response bias. Table 3 displays the probabilities of "yes" responses for each measure along with signal detection parameter estimates providing standardized estimates of discriminability (*d*') and bias (*c*). In the signal detection analyses, the items in the Correction condition were treated as "old," and the items in the Control condition were treated as "new." We included Control headlines, and excluded Repetition headlines, because Control headlines did not include features from Phase 1 and were thus more novel.

Table 3. Probabilities of Correction Classification and Signal Detection Parameter Estimates: Experiment 1

	Classified as Correction ("Yes" Response)			Signal Detection Parameter Estimates	
Phase	Repetition	Control	Correction	*d'*	*c*
Phase 2	.16 [.13, .18]	.03 [.02, .05]	.84 [.81, .87]	2.70 [2.48, 2.93]	.35 [.24, .45]
Phase 3	.15 [.12, .17]	.08 [.06, .10]	.72 [.68, .75]	2.05 [1.83, 2.28]	.41 [.30, .51]

Note. 95% confidence intervals are displayed in brackets.

Awareness of and memory for corrections was first assessed by comparing the extent to which participants indicated that "yes" a headline topic had been (Phase 2) or was (Phase 3) a correction (Table 3, left panel) using a model including Headline Type and Phase (Phase 2 vs. Phase 3) as factors. The model indicated a significant effect of Headline Type, $\chi^2(2) = 1236.54$, $p < .001$, showing that probabilities were significantly higher for Correction than Repetition and Control headlines, smallest *z* ratio = 29.80, $p < .001$, and for Repetition than Control headlines, *z* ratio = 8.59, $p < .001$. There was also a significant effect of Phase, $\chi^2(1) = 8.71$, $p < .01$, that was qualified by a significant interaction, $\chi^2(2) = 37.93$, $p < .001$. The response probabilities for Correction headlines were significantly higher in Phase 2 than in Phase 3, *z* ratio = 5.71, $p < .001$, for Control headlines were significantly higher in Phase 3 than in Phase 2, *z* ratio = 3.70, $p < .001$, and for Repetition headlines were not significantly different between phases, *z* ratio = 0.62, $p = .54$. Together, these results suggest that participants' ability to discriminate headlines that were corrected from the other headline types diminished from Phase 2 to Phase 3. This discrimination difference was verified by fitting separate models to estimates for each signal detection parameter (Table 3, right panel). The model for *d'* confirmed that discrimination was significantly higher during Phase 2 that in Phase 3, $t(94) = 4.00$, $p < .001$, whereas the model for c indicated no significant difference in response bias between phases, $t(94) = 0.78$, $p = .44$.

Next, we partitioned Correction headlines into three groups based on how participants classified them in Phase 2 and Phase 3 (Table 4). This allowed us to (1) assess participants' recall of fake news headline details and (2) compute the proportions of observations comprising each cell in the conditional cued recall analyses (Figure 8, colored points). *Classified as Correction + Fake News Recalled* refers to Correction headlines that were both classified as such (i.e., with a "yes" response) and for which fake news from Phase 1 was correctly recalled after the classification judgment. *Classified as Correction + Fake News Not Recalled* refers to Correction headlines that were classified as such and for which fake news from Phase 1 was *not* subsequently recalled. *Not Classified as Correction* refers to Correction headlines that were not classified as such (i.e., with a "no" response). Since the observations across these three cells were not independent, we only statistically compared the probabilities of fake news recall between the two phases (Table 4, left column) because those responses were of primary theoretical interest. The model indicated significantly higher fake new recall in Phase 2 than Phase 3, $\chi^2(1) = 38.66$, $p < .001$. We did not compare fake news recall in the Repetition and Control headline conditions because participants rarely reported headline details that had not appeared earlier in the experiment ($M \leq .03$).

Table 4. Correction Classification Type Probabilities: Experiment 1

Phase	Correction Classification Type		
	Classified as Correction + Fake News Recalled	Classified as Correction + Fake News Not Recalled	Not Classified as Correction
Phase 2	.64 [.61, .68]	.20 [.17, .23]	.16 [.13, .19]
Phase 3	.50 [.47, .54]	.22 [.19, .25]	.28 [.25, .32]

Note. 95% confidence intervals are displayed in brackets. The classifications above all pertain to Correction headline types. "Classified as Correction + Fake News Recalled" were instances when participants indicated headlines that headline topics were associated with corrections and could recall the fake news details. "Classified as Correction + Fake News Not Recalled" were instances when participants indicated headlines that headline topics were associated with corrections but could not recall the fake news details. "Not Classified as Correction" were instances when participants indicated that headlines were not associated with corrections.

Cued Recall (in Phase 3) Conditioned on Correction Classifications and Fake News Recall (from Phases 2 and 3)

Next, we examined how proactive effects of memory on cued recall accuracy in Phase 3 varied depending on whether corrections that were detected in Phase 2 and accompanied with the retrieval of fake news details were subsequently recalled as such in Phase 3. In the following analyses, we first examined associations between classifications made in Phase 2 and recall accuracy in Phase 3 with models including three levels of Classifications (Classified as Correction + Fake News Recalled, Classified as Correction + Fake News Not Recalled, and Not

Classified as Correction) fitted to real news correct recall and fake news intrusions. We then examined how three levels of Classifications in Phase 3 for only when fake news was recalled in Phase 2 (Classified as Correction + Fake News Recalled) was associated with correct recall and intrusions using separate models for each. We conditioned the analyses on successful fake news recall during Phase 2 because those instances are of primary theoretical interest regarding whether retrieving earlier-studied headlines, would be associated with improved or impaired memory updating for real news headlines. Based on results from similar paradigms (e.g., Putnam et al., 2014; Wahlheim & Jacoby, 2013), we expected that recalling fake news details in Phase 2 would be associated with 1) facilitation in memory for real news details when fake news could subsequently be recollected and 2) interference when fake news could not be subsequently recollected.

Real News Correct Recall

Figure 8A (colored points) displays the conditional probabilities of real news correct recall in Phase 3. A model including Classification in Phase 2 (Classified as Correction + Fake News Recalled vs. Classified as Correction + Fake News Not Recalled vs. Not Classified as Correction) as a factor indicated a significant effect, $\chi^2(2) = 28.82$, $p < .001$, showing that correct recall was significantly higher for instances where corrections were classified as such and fake news was recalled (left green point) than when it was not recalled (left blue and red point), smallest z ratio = 3.59, $p < .01$. There was no significant difference in correct recall in the two cells for which fake news was not recalled in Phase 2, z ratio = 0.96, $p = .60$. Further, a model including the three types of Classifications in Phase 3 for when fake news was recalled in Phase 2 indicated a significant effect, $\chi^2(2) = 90.77$, $p < .001$, showing that recalling fake news was associated with enhanced recall of real news (right green point) compared to when fake news was not recalled in the two other cells (right blue point and red point), smallest z ratio = 7.56, $p < .001$. There was no significant difference in real news recall in the two cells for which fake news was not recalled in Phase 3, z ratio = 1.92, $p = .13$.

To assess proactive effects of memory, correct recall was compared between these Phase 3 conditional cells and Control headlines. Recalling fake news was associated with proactive facilitation, as correct recall was significantly higher than for Control headlines, z ratio = 9.21, $p < .001$. Collapsing across the other conditional cells for which fake news was not recalled, failing to recall fake news was associated with proactive interference, as correct recall was significantly lower than for Control headlines, z ratio = 5.52, $p < .001$. These results showed that overall recall of corrections reflected a mixture of facilitation and interference for which the balance depended on the extent to which participants engaged recollection-based retrieval.

Fake News Intrusion Errors

Figure 8B (colored points) displays conditional probabilities of fake news intrusion errors in Phase 3. The model fitted to intrusions conditionalized on the three Classifications in Phase 2 indicated a significant effect, $\chi^2(2) = 9.03$, $p = .01$, showing that intrusions when fake news was recalled (left green point) than when corrections were classified as such, but fake news was not recalled (left blue point), z ratio = 3.00, $p < .01$. There was no significant difference in intrusions for when corrections were not classified as such in Phase 2 (left red point) compared to the other classifications (left green and blue points), largest z ratio = 2.25, $p = .06$.

The proportions of intrusions for only fake news recall in Phase 2 were then examined based on whether fake news was not recalled in Phase 3 (right blue point and red point). Intrusions for which fake news was recalled were not analyzed because those errors occurred when participants reported the fake news twice in succession, once as being the correct headline detail and once as being the fake news headline detail. These instances of guessing were rare (1%), but the cell appears in Figure 8B to provide a complete picture of how the various response combinations comprised the overall intrusion probability for when fake news was recalled in Phase 2. The model for intrusions for when fake news details were recalled in Phase 2 conditionalized on whether they were not subsequently recalled (right red and blue points) indicated no significant difference, $\chi^2(1) = 0.62$, $p = .43$. These instances were then collapsed and compared with Control headlines to examine proactive effects of memory when fake news was not recalled. The model indicated that average intrusion rate of these cells was significantly higher than the baseline rate for Control headlines, z ratio = 10.79, $p < .001$. This is consistent with the results from correct recall showing that failing to recall fake news was associated proactive interference in memory for updated news headlines, which is a form of familiarity backfire.

Discussion

The results from Experiment 1 showed that correcting fake news increased correct recall for real news relative to when real news headlines appeared once in Phase 2 (i.e., proactive facilitation). These results are inconsistent with familiarity backfire accounts and are generally consistent with conflict saliency accounts. However, conditional analyses revealed evidence for both familiarity backfire and enhanced updating from conflict saliency that depended on whether participants could recall fake news and that it was corrected. These results extend those of Wahlheim et al. (2020) and are consistent with predictions from the MFC framework in showing a mixture of proactive facilitation when earlier-fake news details were retrieved during corrections and later recollected, and interference when earlier-fake news details were retrieved during corrections but not recollected. A similar finding was observed for fake news intrusion errors, showing that recalling earlier-fake news details during corrections but not recollecting corrections was associated with proactive interference. Collectively, these findings suggest that recollection-based retrieval opposed misinformation familiarity, which was increased when corrections were initially detected. The negative effects of recalling fake news details when detecting corrections shown in the absence of recalling those details as such in a subsequent phase joins earlier results in showing that corrections can lead to familiarity backfire (Lewandowsky et al., 2012), but points out the key qualification that such backfire only occurred in specific instances. Thus, neither the conflict saliency nor familiarity backfire account alone could explain the effects of recalling fake news details when detecting corrections on subsequent memory for real news headlines.

Experiment 2

Experiment 1 provided the first characterization of proactive effects of memory when fake news was retrieved during corrections and subsequently recollected on a final memory test. The patterns showing that successful fake news retrieval during corrections was associated with proactive facilitation when fake news was recollected and proactive interference when fake news was not recollected generalizes earlier findings from similar paradigms using different stimulus materials (for a review, see Wahlheim et al., 2021). The primary aim of Experiment 2 was to examine the stability of those associations by testing whether they would replicate using a paradigm variant with an updated material set and slightly different experimental conditions. An exploratory aim of Experiment 2 was to determine how a manipulation intended to affect encoding of fake news details during Phase 1 would subsequently affect the various memory measures in later phases. Since misconceptions about the veracity of everyday news content can be influenced by shared beliefs within social groups (Chan et al., 2017), we examined if varying peer agreement about the veracity of Phase 1 headlines would affect memory for headlines details and the ability to identify and remember fake news being corrected.

Participants were told that while reading each news headline in Phase 1, they should consider if it was true. Then, they were told that most members of a fictional peer group either believed or disbelieved the headline. This trial structure was intended to induce varying levels of congruence between peer and participant beliefs in a somewhat intermixed fashion. Belief congruence was assessed by requiring participants to rate their belief in each headline while a peer-belief message appeared. Based on work showing that belief incongruent information captures attention (Munnich et al., 2007; Vlasceanu et al., 2021) and is more memorable than belief congruent information when the two are directed contrasted (e.g., Maier & Richter, 2013), which is inherent in the Experiment 2 procedure, we predicted that participants would attend more effectively to belief-incongruent headlines, resulting in better memory for fake news details shown on subsequent recall measures in Phases 2 and 3. If so, this should in turn improve memory for reals news, which was shown in Experiment 1 to be positively associated with correct recall of fake news.

Method

Participants

The sample consisted of 76 UNCG undergraduates (53 women, 23 men) ages 18-27 (*M* = 19.05, *SD* = 1.66). As in Experiment 1, our stopping rule was to test as many participants as possible during one semester with the final sample size being a multiple of the total number of experimental formats, which was four (see below). A sensitivity analysis was conducted to estimate the observed power for detecting the smallest effect of interest, which was the difference in real news recall between the Control and Correction [Peers-Believe] condition. The present experiment had 99.50% power to detect an *OR* = 1.55 (see SI 8.2).

Design

The experiment used a 4 (Headline Type: Repetition vs. Control vs. Correction [Peers-Believe] vs. Correction [Peers-Disbelieve]) within-participants design.

Materials

The material set included 60 headline pairs comprising one fake news headline and its correction. This set included the original 41 items corresponding to current events from Experiment 1 and 19 new items about other current events. Based on an item analysis of the Experiment 1 data, we also clarified the wording of three items from the Experiment 1 set. The materials are available on the OSF (https://osf.io/xnvrj/). For counterbalancing, headlines were rotated through conditions and appeared equally often in each condition across participants, as in Experiment 1. Phase 1 included 45 headlines (30 fake news; 15 real news). Phase 2 included 60 headlines [15 real news headlines repeated from Phase 1 (Repetition); 15 real news headlines that corrected Phase 1 fake news appearing earlier with an indicator showing that most peers *believed* the headline (Correction [Peers-Believe]); 15 real news headlines that corrected Phase 1 fake news appearing earlier with an indicator showing that most peers *disbelieved* the headline (Correction [Peers-Disbelieve]); and 15 real news headlines that only appeared in Phase 2 (Control)]. Phase 3 included 60 questions that queried the critical details of Phase 2 headlines.

Procedure

Stimuli were presented electronically using E-Prime 3.0 software (Psychology Software Tools, 2016). The general procedure was like Experiment 1 in many aspects, but there were some key differences regarding Phase 1 (see Figure 7). Prior to Phase 1, participants were told that their tasks would be to study headlines for an upcoming memory test and to silently consider whether they believed each headline was true. This aspect of the instructions was intended to induce mental models about headline veracity that could be affirmed or challenged by peer beliefs. Participants were told that data from a group of previous participants' judgments of headline truth would appear on a number counter that they would view while indicating their own belief. Participants were told the following cover story.

The experiment in which you are about to participate is part of a multi-university study investigating the effects of misinformation and the influence of moderators such as social media. As a part of this collaboration, you will be contributing to a database of responses judging whether information is correct or not. When prompted, select whether you believe the statement to be true or false; you will see the live counter onscreen update to record your response. Thank you for your contribution.

In Phase 1, each headline appeared in a random order for 8000 ms in white font against a black background. A prompt then appeared below the headline text for 5000 ms asking participants to indicate their belief in the headline. Participants made their responses by pressing on a "thumbs up" or a "thumbs down" emoticon, which represented 'believe' and 'do not believe' responses, respectively. A number counter appeared simultaneously below each button displaying the number of fictional peers who believed or disbelieved a headline. The integer shown on each counter was selected randomly for each trial from a range of percentages from a fictional sample of 1100 peers. In the Peers-Believe condition, the 'believe' response counter showed an integer ranging from 65-75% of the total sample (e.g., 770), while the 'do not believe' counter displayed an integer representing the remaining percentage (e.g., 330). In the Peers-Disbelieve condition, this proportion was reversed between counters.

We selected moderate values instead of more extreme values because more extreme values seemed less plausible from a participant's perspective. However, we acknowledge that social influence can occur when extreme values are used (Kim, 2018; Vlasceanu & Coman, 2021). For Repetition headlines, the percentages ranged from 48-52% so that the presence of counters was not confounded with headline veracity. The prompt color changed from white to yellow and the counter updated when participants entered their ratings and remained on the screen for the entire 5000 ms trial duration. After Phase 1, participants completed a math distractor task for 10 minutes, indicating whether solutions to simple addition problems were even or odd numbers. This distractor task was included to lower the change detection rate from Experiment 1 (M = .87) to provide a more even distribution of items among cells for the recall analyses conditioned on Phase 2 correction detection responses. This modification was especially useful for examining differences in fake news intrusion errors (see below).

Phase 2 was the same as in Experiment 1. A mixed-effect model with Participants and Items as random intercept effects indicated that the estimated marginal mean of yes/no durations was significantly lower for Repetition (M = 1477 ms, 95% CI = [1254, 1700 ms]) than Correction [Peers-Disbelieve] (M = 1770 ms, 95% CI = [1547, 1993 ms]) headlines, z ratio = 2.79, p = .03. The estimate for Control headlines (M = 1741 ms, *95% CI* = [1518, 1964 ms]) was not significantly different than the estimates for Repetition, Correction [Peers-Believe] (M = 1620 ms, 95% CI = [1397, 1843 ms]), Correction [Peers-Disbelieve] headlines, largest z ratio = 2.52, p = .06. Finally, there was no significant duration difference between the Correction [Peers-Believe] and Correction [Peers-Disbelieve] headlines, z ratio = 1.42, p = .49.

Results

Cued Recall Performance (Phase 3)

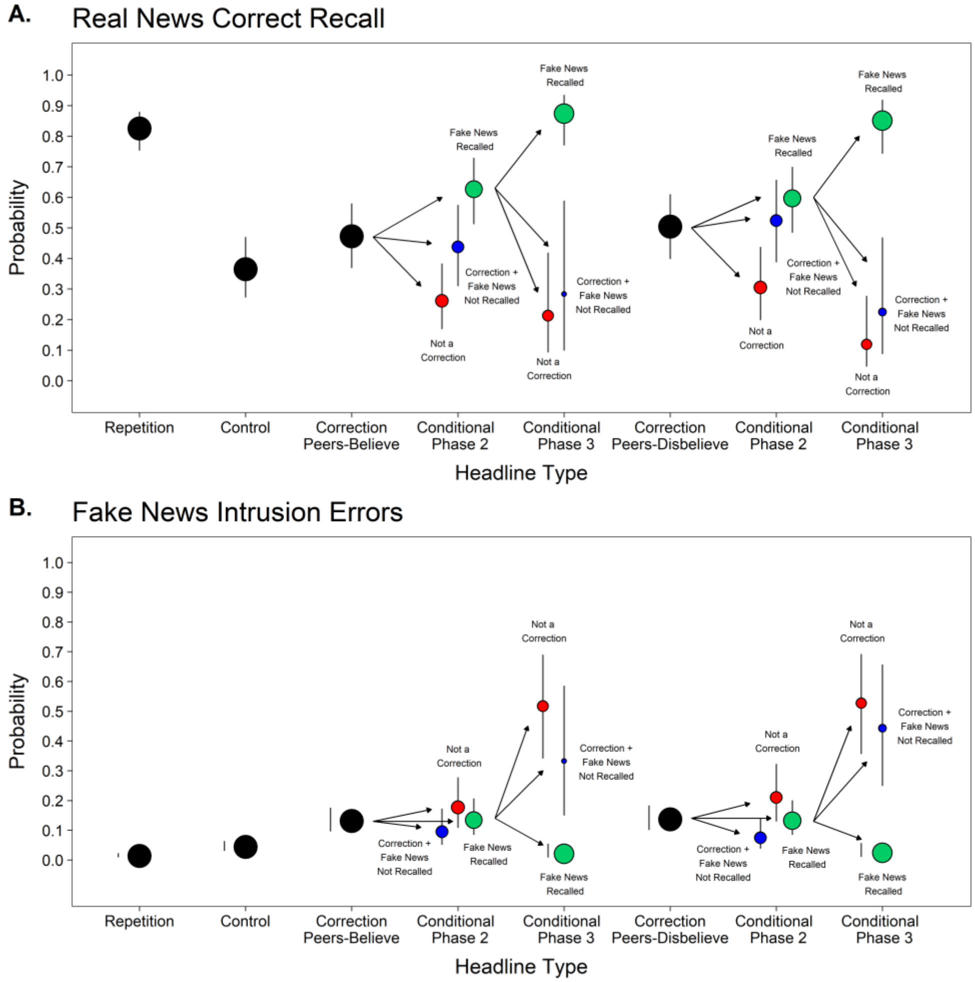
Real News Correct Recall

Figure 9A (black points) displays the overall probabilities of real news correct recall in Phase 3. A model including Headline Type as a factor indicated a significant effect, $\chi^2(3) = 364.47$, $p < .001$. Correct recall was significantly greater for Repetition than the other three headline types, smallest *z* ratio = 13.78, $p < .001$. Correct recall was also significantly higher for Correction [Peers-Believe] and Correction [Peers-Disbelieve] headlines than Control headlines, smallest *z* ratio = 4.35, $p < .001$. There was no significant difference between the Correction [Peers-Believe] and Correction [Peers-Disbelieve] headlines, *z* ratio = 1.25, $p = .59$. Replicating Experiment 1, these results indicating overall proactive facilitation in memory for real news that corrected fake news, and thus, no overall familiarity backfire effect.

Fake News Intrusion Errors

Figure 9B (black points) displays the overall probabilities of fake news intrusion errors in Phase 3. A model including Headline Type as a factor indicated a significant effect, $\chi^2(3) = 204.22$, $p < .001$. Intrusions were significantly higher for both types of Correction headlines than the other headline types, smallest *z* ratio = 8.21, $p < .001$. There were no significant differences between Correction [Peers-Believe] and Correction [Peers-Disbelieve] headlines, *z* ratio = 0.45, $p = .97$, but intrusions were significantly higher for Control headlines than Repetition headlines, *z* ratio = 5.40, $p < .001$. Replicating Experiment 1, these findings show that fake news exposure in Phase 1 led to proactive interference when participants attempted to recall of real news details from Phase 2.

Figure 9. Recall of Real News and Intrusions of Fake News: Experiment 2



Note. Probabilities of real news correct recall (Panel A) and fake news intrusion errors (Panel B) as a function of Headline Type in Experiment 2. Black points represent probabilities for all observations. Colored points represent probabilities conditioned on correction classification types in Phase 2 and 3. The cells represent corrections that were classified as such and for which fake news was recalled (green points), corrections that were classified as such and for which fake news was not recalled (blue points), and corrections that were not classified as such (red points). The size of each point indicates the relative proportion of observations in each cell. Error bars are 95% confidence intervals and are displayed adjacent to the points when the intervals lengths are shorter than point diameters.

Correction Classifications and Fake News Recall (Phases 2 and 3)

Following our approach in Experiment 1, we identified the role of detection of fake news corrections and recollecting that fake news had been corrected in cued recall accuracy by first assessing awareness of and memory for corrections in Phase 2 and in Phase 3 (Table 5). The extent to which participants indicated that "yes" a headline topic was or had been a correction (left panel) was compared between phases using a model including Headline Type and Phase as factors. The model indicated significant effects of Headline Type, $\chi^2(3) = 2255.02$, $p < .001$, showing that probabilities were significantly higher for Correction [Peers-Believe] and Correction [Peers-Disbelieve] than Repetition and Control headlines, smallest z ratio = 34.96, $p < .001$. There was no significant difference between the Correction [Peers-Believe] and Correction [Peers-Disbelieve] headlines, z ratio = 0.71, $p = .89$, or Repetition and Control headlines, z ratio = 0.83, $p = .84$. There was also a significant effect of Phase, $\chi^2(1) = 99.55$, $p < .001$, showing that probabilities were significantly higher in Phase 2 than in Phase 3. There was a significant interaction, $\chi^2(3) = 48.80$, $p < .001$, showing that the probabilities for Correction [Peers-Believe], Correction [Peers-Disbelieve], and Control headlines were significantly higher in Phase 2 than in Phase 3, smallest z ratio = 3.47, $p < .001$, and the probabilities for Repetition headlines were not significantly different between phases, z ratio = 1.87, $p = .06$. Collectively, these findings show that participants were less able to discriminate Correction headlines from other headline types in Phase 3 than in Phase 2.

Table 5. Probabilities of Correction Classification and Signal Detection Parameter Estimates: Experiment 2

					Signal Detection Parameter Estimates			
	Classified as Correction ("Yes" Response)				*d'*		*c*	
Phase	Repetition	Control	Correction [Peers-Believe]	Correction [Peers-Disbelieve]	Correction [Peers-Believe]	Correction [Peers-Disbelieve]	Correction [Peers-Believe]	Correction [Peers-Disbelieve]
Phase 2	.11 [.09, .13]	.14 [.12, .16]	.73 [71, .76]	.74 [.71, .76]	1.90 [1.68, 2.13]	1.89 [1.66, 2.11]	.26 [.15, .36]	.26 [.15, .37]
Phase 3	.13 [.11, .15]	.09 [.08, .11]	.58 [.55, .61]	.59 [.56, .62]	1.61 [1.39, 1.84]	1.63 [1.41, 1.86]	.58 [.47, .69]	.57 [.46, .68]

Note. 95% confidence intervals are displayed in brackets.

To verify this claim, separate models were fitted to the estimates from each signal detection parameter (right panel). The model for *d'* indicated that discrimination was significantly higher during Phase 2 than in Phase 3, $F(1, 300) = 5.74$, $p = .02$. In addition, the model for *c* indicated that estimates were significantly higher in Phase 2 than in Phase 3, $F(1, 300) = 33.36$, $p < .001$, showing that participants became more conservative across phases. The less conservative responding in Phase 2, which did not replicate results from the first experiment, may have occurred because corrections comprised half of the items here and only a third of the items in the first experiment. No other effects were significant, *largest* $F(1, 300) = 0.02$, $p = .88$.

Also following our approach in Experiment 1, we next partitioned the classifications for Correction headlines in each phase into three types (Table 6) to assess the extent to which participants recalled fake news and to compute the proportions of observations comprising cells in the conditional analyses of cued recall (Figure 9, colored points). A model including all headline conditions indicated no significant difference in fake news recall between the two Correction headline conditions, z ratio = 0.44, $p = .97$, and that fake news recall for those conditions was significantly higher in Phase 2 than Phase 3, smallest z ratio = 5.59, $p < .001$. Participants rarely reported details from the fake news headlines that had not appeared in the experiment for Repetition and Control headlines ($M \leq .02$).

Table 6. Correction Classifications Type Probabilities: Experiment 2

	Correction Classification Categories Probabilities					
	Classified as Correction + Fake News Recalled		Classified as Correction + Fake News Not Recalled		Not Classified as Correction	
Phase	Correction [Peers-Believe]	Correction [Peers-Disbelieve]	Correction [Peers-Believe]	Correction [Peers-Disbelieve]	Correction [Peers-Believe]	Correction [Peers-Disbelieve]
Phase 2	.50 [.48, .53]	.51 [.48, .54]	.24 [.20, .27]	.23 [.20, .27]	.26 [.22, .29]	.26 [.23, .30]
Phase 3	.41 [.38, .44]	.41 [.39, .44]	.18 [.15, .21]	.18 [.14, .21]	.42 [.38, .46]	.42 [.38, .46]

Note. 95% confidence intervals are displayed in brackets. The classifications above all pertain to Correction headline types. Classified as Correction + Fake News Recalled were instances when headlines were classified as corrections and participants could recall the fake news details. Classified as Correction + Fake News Not Recalled were instances when headlines were classified as corrections, but participants could not recall the fake news details. Not Classified as Correction were instances when headlines were not classified as corrections.

Cued Recall (in Phase 3) Conditioned on Correction Classifications and Fake News Recall (from Phases 2 and 3)

The role of detecting corrections and recalling fake news on the proactive effects of memory was examined using the same approach as in Experiment 1.

Real News Correct Recall

Figure 9A (colored point) displays the conditional probabilities of real news correct recall in Phase 3. A model including Classification in Phase 2 and Headline Type as factors indicated a significant effect of Classification, $\chi^2(2) = 41.55$, $p < .001$, no significant effect of Headline Type, $\chi^2(1) = 0.25$, $p = .62$, and no significant interaction, $\chi^2(2) = 1.81$, $p = .40$. Pairwise comparisons showed that correct recall was significantly higher when corrections were classified as such and fake news was recalled (left green point) than the other two classifications (left blue and red point), smallest z ratio = 2.52, $p = .03$. Correct recall was also significantly higher when corrections were classified as such (left blue point) than when they were not (left red point), z ratio = 3.59, $p = .001$.

Further, a model including the three types of Classifications in Phase 3 when fake news was recalled in Phase 2 and Headline Type as factors indicated a significant effect of Classification, $\chi^2(2) = 80.45$, $p < .001$, no significant effect of Headline Type, $\chi^2(1) = 1.24$, $p = .27$, and no significant interaction, $\chi^2(2) = 0.48$, $p = .79$. Pairwise comparisons showed that correct recall of real news was significantly higher when fake news was recalled (right green points) than when corrections were classified as such and fake news was not recalled (right blue points) and when corrections were not classified as such (right red points), smallest z ratio =

6.03, $p < .001$. There was no significant difference in correct recall in the two cells for which fake news was not recalled (right blue and red points) z ratio = 1.10, $p = .52$.

To examine the proactive effects of fake news exposure on subsequent memory for corrections, we first collapsed the Correction [Peers-Believe] and Correction [Peers-Disbelieve] conditions into a single Correction condition because there were no differences in conditional recall between these two conditions. We then compared conditional correct recall in the Correction headline conditions with the Control condition. A model including Headline Type as a factor with the three classification types for corrections in Phase 3 when fake news was recalled in Phase 2 and Control headlines as the levels indicated that recalling fake news was associated with proactive facilitation, as correct recall was significantly higher for those Correction than Control headlines, t ratio = 12.04, $p < .001$. Collapsing across the other conditional cells for which fake news was not recalled, failing to recall fake news was associated with proactive interference, as correct recall was significantly lower than for those Correction than Control headlines, t ratio = 3.88, $p < .001$. Replicating Experiment 1, these results show that overall recall of corrections reflected a mixture of facilitation and interference for which the balance depended on the extent to which participants engaged recollection-based retrieval. However, a novel finding here was that higher memory accuracy for real news when corrections were classified as such but no fake news details were retrieved relative to not classifying a correction.

Fake News Intrusion Errors

Figure 9B (colored points) displays conditional fake news intrusion errors in Phase 3. A model including Classification in Phase 2 and Headline Type as factors indicated a significant effect of Classification, $\chi^2(2) = 11.65$, $p < .01$, showing that there were fewer intrusion errors for when corrections were classified and fake news was not recalled (left blue points) than when corrections were not classified as such (left red points), *z* ratio = 3.43, $p < .01$. There was no significant difference in intrusions for when corrections were classified as such and fake news was recalled (left green points) compared to the other classifications (left blue and red points), largest *z* ratio = 1.93, $p = .13$. There was no significant effect of Headline Type, $\chi^2(1) < 0.01$, $p = .97$, and no significant interaction, $\chi^2(2) = 0.92$, $p = .63$.

The model for intrusions for when fake news details were recalled in Phase 2 conditioned on whether they were not subsequently recalled (right red and blue points) indicated no significant effect of Classification, $\chi^2(2) = 2.20$, $p = .14$, no significant effect of Headline Type, $\chi^2(1) = 0.13$, $p = .72$, and no significant interaction, $\chi^2(1) = 0.20$, $p = .66$. These instances were then collapsed and compared with Control headlines to examine proactive effects of memory when fake news was not recalled. The model indicated that average intrusion rate of these cells was significantly higher than the baseline intrusion rate for Control headlines, *t* ratio = 18.14, $p < .001$. As in Experiment 1, the overall patterns here are consistent with the results from correct recall showing that failing to subsequently recall fake news was associated proactive interference in memory for updated news headlines, which is a form of familiarity backfire. However, the fewer intrusions associated with classifying a correction but not recalling fake news is a novel finding in this study.

Peer Beliefs (Phase 1) and Memory for Fake News Corrections (Phase 2 and 3)

We also explored if the congruence between peer and participant beliefs influenced the encoding of fake news headlines in Phase 1 and its consequences for performance on memory measures in subsequent phases (for a complete description of these analyses, see SI 10). A manipulation check revealed that peer belief ratings influenced participant belief ratings in Phase

1: fake news headlines were rated as more believable when participants were told that most of their fictional peers also believed rather than disbelieved those headlines (SI 10.1). However, contrary to our hypothesis that mismatching peer and participant beliefs would improve memory for fake and real news, there was no difference in fake news recall during Phase 2 and Phase 3 or in real news recall in Phase 3 depending on whether peer and participant beliefs matched or mismatched in Phase 1 (SI 10.2).

Discussion

Replicating Experiment 1, recall of real news was better when it corrected fake news than when it appeared only once during Phase 2. The overall enhancement in memory for real news resulting from it correcting fake news reflected a mixture of proactive facilitation and interference that depended on the extent to which fake news was retrieved during corrections and subsequently recollected. As in Experiment 1, recollecting earlier-retrieved fake news was associated with enhanced memory updating, whereas not recollecting earlier-retrieved fake news was associated with interference driven by familiarity backfire. This complex interplay could not be accounted for by familiarity backfire or conflict saliency accounts alone. However, these findings are consistent with predictions from the MFC framework that subsumes the key mechanisms proposed by those accounts and proposes a moderating role for recollection-based retrieval. These findings build on Wahlheim et al. (2020) by showing that even without overt reminders of fake news headlines, retrieval of fake news leads to facilitation or interference in real news recall depending on if recollection-based retrieval is engaged on the final recall test. Finally, the manipulation of peer and participant beliefs showed no effects on misinformation accessibility, which was inconsistent with our hypothesis that belief incongruent information would be better encoded and retrieved.

General Discussion

Research on misinformation corrections has shown that exposure to everyday misinformation has a continued influence on memories upon which beliefs are based (for a review, see Ecker et al., 2022). Given the current dilemma posed by the high prevalence of fake news in media outlets, determining ways to counteract its effects on memory accuracy is a priority. Despite the urgency of this issue, such effects are not yet well-established. The primary goal of the present study was therefore to examine the role of recalling fake news details when initially detecting corrections and when attempting to subsequently recall the real news details that had corrected the fake news. The current experiments provided key data points for evaluating leading accounts of how fake news proactively affects memory for accurate information of the sort presented by news outlets. The present findings showed that retrieving fake news details during corrections can improve memory for real news headlines when fake news is later recollected and impair memory for real news headlines when fake news is not later recollected. The familiarity backfire and integrative encoding views could both partly account for different aspects of these findings, but the MFC framework better accounted for the complete pattern by subsuming key assumptions of those views and including a moderating role for recollection-based retrieval. In what follows, we discuss the implications of these findings for leading theoretical views and applications in everyday life.

Familiarity Backfire and Recollecting Detected Corrections

Fake news corrections had once been discouraged because of concerns that they increase the fluency of misinformation by triggering its retrieval (Lewandowsky et al., 2012), but subsequent studies motivated a revision of that suggestion because corrections more often improve memory and belief accuracy (Lewandowsky et al., 2020). In fact, a recent review of the backfire effect literature (Swire-Thompson et al., 2020) concluded that belief-based familiarity backfire effects are not robust and are often an artifact of design characteristics that promote regression of beliefs to the mean, such as unreliable measures and pre-post designs without a control group. Taken with the established finding that repeating headlines increases beliefs when corrections are absent (e.g., Fazio et al., 2019), Swire- Thompson et al. (2020) suggest pairing corrective information clearly with original misinformation, presumably to promote associative memory for details and their veracity and counteract erroneous fluency-based beliefs. However, this approach cannot always be achieved in the wild, especially when the rapid availability of media-based misinformation outpaces the release of fact-checked content, thus leaving people responsible for self-detecting corrections and deciding whether to engage retrieval mode to recall the original fake news details.

Although the present study did not directly examine how corrections influence beliefs, the findings did improve our understanding of when self-identified corrections improve or impair the memories upon which beliefs can be based. Consistent with prior findings (for a review, see Lewandowsky et al., 2020), the present results suggest that fake news corrections improved memory for updated headlines when misinformation had become associated with corrective information well enough to support recollection-based retrieval. These findings contribute to the backfire literature by highlighting the need to consider how identifying corrections by using their features to cue retrieval of misinformation can determine the extent to which backfire effects on memory occur *within* participants. This within-participant method of conditioning real news recall reveals how the influences of opposing processes are obscured when only analyses of aggregated data are considered. We therefore recommend that research aimed at identifying optimal correction formats leverage the conditional techniques used here and in related memory updating studies (e.g., Putnam et al., 2014; Jacoby et al., 2015). This approach could be useful in designing interventions that optimize participant strategies and correction details to promote misinformation retrieval and its associative encoding with corrective information.

Integrative Encoding of Misinformation and Corrections

The suggestion that associative encoding of misinformation and corrections can improve memory accuracy in the service of counteracting erroneous beliefs is central to integrative encoding accounts of memory and belief updating (e.g., Kendeou et al., 2014; Ecker et al., 2017; Wahlheim & Jacoby, 2013). According to those accounts, memory and belief updating are enabled when misinformation and correct information from two events becomes encoded together with a representation that identifies which details are true. Evidence for integrative encoding in memory paradigms can be inferred from response dependencies showing that updated information is more likely to be recalled accurately when original information is also recalled (for a review, see Wahlheim et al., 2021). The present finding showing that recall for real news was more accurate when earlier-retrieved fake news was subsequently recollected is compatible with an integrative encoding account of the proactive facilitation.

The present findings also replicate results showing similar facilitation in conditional recall of fake news headlines (Wahlheim et al., 2020) and build on those results by specifying the role of retrieving fake news details while encoding corrections. The present findings were also the first to show proactive facilitation in real news corrections in aggregate recall when participants were tasked with identifying corrections and retrieving fake news details on their own. This finding diverges from earlier results showing that corrections appearing without fake news details were remembered as well as real news that had not corrected fake news (Wahlheim et al., 2020). This discrepancy may have arisen because the correction detection procedure here that required generating prior-list fake news details encouraged participants to engage in study-phase retrievals more often, thus promoting integrative encoding for more headlines. This account is reminiscent of earlier findings showing that looking back to earlier information when encoding changes leads to proactive facilitation in recall of recently learned associations (Jacoby et al., 2015) via integrative encoding.

More broadly, the premise that integrative encoding facilitates updating also aligns with work in the educational literature that emphasizes the benefits of refutation texts on knowledge revision (Kendeou et al., 2014, 2019). These studies suggest that knowledge revision is successful when the refuted information is co-activated and integrated with the newly encoded correct information in memory. The present work broadens our knowledge of integrative encoding by illuminating its effects on memory updating in the context of fake news. Consistent with recent studies (Brashier et al., 2021; Grady et al., 2021), we used real-world news headlines that were corrected by fact-check verified headlines to enhance the applicability of our findings. Importantly, our analytic approach advances theoretical proposals in related work by identifying when integrative encoding was not effective and fake news produces interference by specifying a role for recollection-based retrieval.

The Role of Recollection-Based Retrieval

The importance of recollection-based retrieval in overcoming interference is inherent in the dual-process account of the CIE (Jacoby, 1991, 1999; for a review, see Lewandowsky et al., 2012), which assumes that misinformation continues to exert its influence when automatic memory is unopposed by strategic recollection. Findings to corroborate this perspective comes from studies showing how susceptibility to the CIE is heightened when recollection is less available, such as with older participants (Skurnik et al., 2005; Swire et al., 2017a), longer retention intervals (Brashier et al., 2021; Walter & Tukachinsky, 2020), and when attention is divided (Ecker et al., 2011). However, a limitation of the dual-process account is that without modification it cannot account for findings showing that recollection of misinformation is associated with correct recall (Moore & Lampinen, 2016) or recent neuroimaging evidence implying that misinformation recollection drives the CIE and not misinformation familiarity (Brydges et al., 2020; Gordon et al., 2019). Together, these observations emphasize the need to consider a comprehensive account of misinformation correction effects that combines standard dual-process and integrative encoding perspectives.

Cohort Agreement

The present study also showed that the congruence between participant and peer beliefs about fake news accuracy in Phase 1 did not lead to differences in recall of fake news details in Phases 2 and 3 or in recall of real news details in Phase 3. This was somewhat surprising given that belief incongruence can induce skepticism, leading to a more analytic evaluation of headlines resulting from prediction errors that upregulate attention (Munnich et al., 2007; Vlasceanu et al., 2021). Although the reason for the absence of a belief congruence effect is unclear, one possibility is that belief congruence and incongruence both improved memory via different routes. For example, belief congruence could have enhanced encoding by improving encoding fluency (Schwarz et al., 2016), memory representations (Levine & Murphy, 1943), and integration with schemas (Pratkanis, 1989). Another possibility is that the values we selected for peer beliefs may have not been extreme enough to induce social influence (cf. Kim, 2018; Vlasceanu & Coman, 2021) that would stimulate elaborative processing when participants consider how their beliefs contradicted others. These possibilities could be examined by varying the extremity of peer beliefs and including a contrast condition for which peer beliefs are not disclosed to participants. Furthermore, the potential effects of belief congruence may have been limited because our materials included fairly partisan-neutral content. Since misinformation that aligns with political ideologies can be resistant to corrections (for a review, see Swire-Thompson et al., 2020), future research could examine how belief congruency interacts with memory while manipulating the alignment of misinformation content with participant partisanship.

Inconsistent Findings in the Present Experiments

While the present experiments converged in identifying the roles of integrative encoding and recollection-based retrieval in memory for corrections of fake news, there were a few unexpected differences in conditional recall performance. Unlike Experiment 1, in Experiment 2, cued recall accuracy in Phase 3 conditioned on classifications in Phase 2, showed that real news recall was higher when corrections were classified as such without fake news recall than those not classified as such and the inverse for fake news intrusions. Furthermore, both experiments showed no differences in real news recall or fake news intrusions when corrections for which fake news was recalled in Phase 2 were classified as such in Phase 3 without fake news recall and when they were not classified as corrections, which had not earlier been shown when using comparable materials (Wahlheim et al., 2020). Although there is no clear explanation, these discrepancies might reflect different bases for classifying headlines as having been corrected, such as partial recall of fake news details. However, a direct test of this supposition is required to determine its plausibility.

On the Relationship between Memory and Beliefs

While prior work has largely focused on the effects of misinformation exposure on beliefs (for a review, see Pennycook & Rand, 2021), there have also been a few studies examining misinformation effects on memory (Ecker et al. 2011; Wahlheim et al. 2020). We believe there is much to be gained from investigating the effects on both of these measures because memory is one primary basis for beliefs (e.g., Berinsky, 2017; Kowalski & Taylor, 2017). Evidence for this assumption comes from studies showing differences in belief regression that coincide with recollection differences. For example, myth beliefs following corrections regressed more to baseline beliefs at longer than shorter retention intervals and for older than younger adults (Swire et al., 2017a). These results imply that when myth corrections were less well remembered, participants endorsed erroneous beliefs more. Taken with the present findings, these differences in belief regression inspire new avenues of inquiry about how beliefs vary based on the various retrieval combinations incorporating fake news recall measures across phases in paradigms such as those used here.

Limitations

The present study had two primary limitations that were essential to note (though one could certainly identify others). First, the self-paced nature of the study phase could have contributed to memory differences for fake news corrections because participants spent slightly more time making judgments for those items in Experiment 1, and the headline remained on the screen until participants typed their recall response. This additional time could have been partly responsible for the proactive facilitation observed in overall real news recall. However, there was likely still a prominent role of integrative encoding in this recall advantage as an interference-based account would assume that more time spent encoding alternative associations when fake news was recalled would lead to subsequent proactive interference in overall recall. Future studies could control for encoding time differences by holding the duration constant across headline types and omitting Phase 1 recall following correction detection judgments. However, the average duration of correction detections was comparable between control and correction headlines in Experiment 2, providing evidence against a differential encoding account of proactive facilitation in overall recall.

Second, the stimulus materials and presentation format of corrections provided ecologically valid information content and a way to assesses the downstream consequences of recalling fake news during corrections. Although consumers may read news headlines as short declarative sentences, as in the present study, headlines sometimes appear with images, and news is sometimes delivered in modalities with unfolding temporal structures, such as videos or podcasts. Additionally, it is not often that an outside source tells someone to read headlines and compare the content to existing memories for earlier-read headlines.

Conclusion

To conclude, this study extends the line of research highlighting the importance of recollection-based retrieval following retrieval of fake-news details during corrections for subsequent updating memory for real news details. It allowed us to test competing predictions from leading theoretical perspectives on memory and belief updating, namely the familiarity-backfire and integrative encoding accounts. The present findings implicated roles for mechanisms from both accounts along with a critical moderating role for recollection-based retrievals. These findings suggest that successfully retrieving fake news details when reading real news headlines can promote the comparisons necessary to encode associations and support later recollection of news headlines and their veracity. Future interventions may be improved by considering how interactions between retrieval strategies engaged during encoding and the inclusion of overlapping features between real and fake news headlines can promote enduring memory representations upon which beliefs can be based.

CHAPTER IV: CORRECTING FAKE NEWS HEADLINES AFTER REPEATED EXPOSURE: MEMORY AND PERCEIVED ACCURACY IN YOUNGER AND OLDER ADULTS

Abstract

Adults of all ages are exposed to varying amounts of fake news. The efficacy of fake news corrections in improving memory and belief accuracy may depend on fake news exposure and consumer age. Two experiments tested the competing predictions that repeating fake news before corrections will impair or improve memory and belief accuracy, especially for older adults. Participants read real and fake news headlines that appeared once or thrice then identified fake news corrections among real news headlines. Recognition and cued recall tests assessed memory for real news, fake news, if corrections occurred, and beliefs in retrieved details. Repeating fake news increased detecting and remembering of corrections, real news memory, and fake news intrusions. Younger adults detected and remembered more corrections and enjoyed greater associated benefits to real news retrieval than older adults. When corrections were not remembered, earlier repetitions of fake news increased memory errors mistaking fake for real news. Earlier repetitions of fake news also affected older adults more negatively when the task was to generate retrievals of real news. Both age groups showed comparable overall belief accuracy and comparably more accurate beliefs for falsely recognized fake news following earlier repetitions of fake news. However, when participants did not remember corrections, older adults showed higher beliefs in fake news intrusions, which are less accurate, than younger adults. These findings indicate that the extent to which fake news repetitions impaired and

improved memory and belief accuracy depended on age differences detecting and remembering corrections.

Significance Statement

Fake news exposure can negatively impact memories and beliefs for all people. To combat such exposure, we must understand how corrections mitigate these effects across age groups, especially younger and older adults. One account proposes that more fake news exposure before corrections should improve discernment between real and fake news for younger but not older adults, whereas another account proposes that both groups should enjoy comparable improvement. We tested these competing predictions here. Participants first read real and fake headlines, with the latter appearing once or three times. Participants then saw all real news headlines before attempting recognize or recall the headlines and indicating their beliefs in remembered details. More exposure to fake news did not impair overall memory or belief accuracy for either age group. In fact, such exposure allowed both groups to better remember how real and fake news details were associated. Memory and belief accuracy benefitted from these associations slightly more for younger than older adults. Importantly, neither group showed stark impairments. The main negative effects of fake news exposure occurred only when participants could not remember that real news corrections had appeared. Collectively, these findings suggest that additional exposure to fake news could improve memory for its veracity and how it relates to real news. This may be helpful for both younger and older adults because it better contrasts true and false details, thus enhancing their discernment.

Introduction

Exposure to fake news stories on the internet can influence beliefs that negatively affect everyday decisions for adults of all ages. Underscoring this, COVID-19 misinformation has been shown to reduce self-reported intentions to vaccinate and follow health guidelines (Loomba et al., 2021; Roozenbeek et al., 2020). Fake news can be distributed rapidly across many internet platforms. People may thus be repeatedly exposed false information before fact-checkers can issue corrections, increasing the potential for misguided decisions. Such exposure may have more detrimental consequences for older than younger adults because older adults experience poorer recollection of contextual details (e.g., Jennings & Jacoby, 1993). This creates a fundamental issue in modern society—we need methods for mitigating the negative effects of fake news exposure that take age-related cognitive differences into account. Accomplishing this requires identifying the mechanisms underlying the effects of repeated misinformation exposure on memory and perceived accuracy in younger and older adults.

Repeated exposure to misinformation before corrections can negatively or positively affect memory for true details and perceptions of accuracy in younger adults. Repeating misinformation during corrections sometimes increases misperceptions of misinformation accuracy (Nyhan et al., 2014). Conversely, such repetitions sometimes diminish the influence of misinformation influence on inferential reasoning (Ecker et al., 2017). Moreover, the negative effects of fake news re-exposure during corrections have been shown to occur more often when people forget that misinformation had been corrected (Wahlheim et al., 2020; Kemp, et al., 2022ab). Although studies have begun to characterize the consequences of repeating misinformation before corrections (Ecker et al., 2011, 2017; Pennycook et al., 2018), no studies have assessed how repeating fake news from internet headlines before fact-check verified corrections affects retrieval of true details and beliefs in those details. This issue needs to be addressed to clarify when fake new exposure before corrections impairs or improves the quality of remembered details that guide beliefs and everyday decisions.

We addressed this issue here in two experiments that varied fake news headline exposure before corrections. We measured detection of corrections with updated real news details and subsequent memory for and beliefs in retrieved details for younger and older adults. Repeated exposure to fake news may be particularly problematic for older adults because their reduced memory for associative information (for a review, see Park & Festini, 2017), such as veracity information, may undermine their ability to update memories and beliefs. Understanding the consequences of repeating fake news before corrections in these age groups is needed because older adults have been shown to engage more with fake news social media platforms (Grinberg et al., 2019; Guess et al., 2019) and perceive fake news as more accurate, especially those that have been repeated (Lyons, 2023) presumably because of their limited digital literacy (Brashier & Schacter, 2020; but see Allcott & Gentzkow, 2017; Pehlivanoglu et al., 2022). We motivate the present study below by summarizing select findings and theories from the literatures on the continued influence of misinformation and age-related differences in memory for associative information.

The consequences of misinformation exposure have traditionally been examined using a narrative-based paradigm in which participants read an unfolding fictitious event (Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988). The event includes a specific misinformation detail that is later corrected or not. Correction efficacy is then assessed with inferential reasoning questions that evaluate the influence of the misinformation detail. A consistent finding is that people continue to rely on the misinformation in their inferential reasoning, even when they remember that a correction was issued earlier (for a review, see Lewandowsky et al., 2012). This continued influence effect is robust as it has been replicated using various materials (e.g., news reports and myths) in laboratory and online settings (Ecker et al., 2011; Rich & Zaragoza, 2016, 2020; Desai & Reimers, 2019). Although corrections vary in their efficacy—being more effective when they are coherent, congruent with existing beliefs, and are provided by credible sources—they do not entirely eliminate the influence of misinformation (for a review and meta-analysis, see Walter & Tukachinsky, 2020).

Several theories have been invoked to account for the continued influence effect (for a review, see Lewandowsky et al., 2012). Most germane to the present study, which concerns memory and belief interactions, are dual-process and source memory theories. We consider other relevant theories in the General Discussion. Dual-process theories propose that retrieval can be based on controlled recollection of contextual information, which can include veracity and source details, or automatic familiarity that reflects memory strength without context (Ayers & Reder, 1998; Jacoby, 1991; Yonelinas, 2002). Accordingly, after misinformation and corrections are encoded, they co-exist in memory and compete for activation at retrieval. Repeating misinformation increases its familiarity and ease of processing (Schwarz et al., 2007), and therefore its potential to influence subsequent memory and reasoning. When contextual details are not recollected, the misinformation made familiar by repetition becomes a more attractive response candidate. Similarly, the source monitoring framework proposes that people can separately remember both content and the source from which it originated (Johnson et al., 1993). Consequently, remembering the content without the source, especially when misinformation is more familiar can lead to memory errors. These views can account for the finding that misinformation repetitions increase the continued influence effect (cf. Ecker et al., 2011).

Repeating misinformation during retractions has also been shown to increase its familiarity to the extent that it backfires by increasing erroneous beliefs in naturalistic tasks other than the narrative-based paradigm. In a study that introduced the myth vs. fact message frame, participants read a flyer juxtaposing myths and facts associated with the flu vaccine (Skurnik et al., 2007; as cited in Schwarz et al., 2007). Beliefs were assessed immediately or after 30 minutes by requiring participants to identify whether statements were myths or facts. Performance was near perfect on an immediate test. But after the delay, participants misremembered many myths as being true and expressed more negative attitudes towards the flu vaccine relative to participants who had not seen the flyer. Consistent with dual-process theories, these findings were attributed to poorer recollection after the delay giving rise to automatic influences of myth familiarity that was increased by presenting myths with facts (also see, Begg et al., 1992; Skurnik et al., 2005). Support for this view has since been shown in studies reporting that repeating misinformation with corrections increases misperceptions of misinformation accuracy (e.g., Autry & Duarte, 2021; Nyhan et al., 2014; Peter & Koch, 2016; Pluviano et al., 2017, 2019). These findings suggest that more exposure to everyday misinformation, such as fake news stories, may increase its familiarity and therefore its potential to proactively interfere with memory for corrective details.

Despite the work suggesting that repeating misinformation creates the risk of backfire, recent research casts doubt of the robustness of such effects. Indeed, a recent meta-analysis indicates that backfire effects are less pervasive than some have assumed and could reflect design artifacts, such as unreliable measurement and insufficiently powered designs (Swire-Thompson et al., 2020). Additionally, studies using various misinformation correction paradigms have shown that reminders can reduce the influence of misinformation. For example, in the narrative-based paradigm, explicit misinformation reminders with corrective details have been shown to reduce misinformation reliance in inferential reasoning (Ecker et al., 2017). This effect was attributed to reminders promoting co-activation of false and true information that improved encoding of the details and their conflict. This view is generally compatible with research suggesting that knowledge revision is improved when conflicting details are detected (Kendeou et al., 2014, 2019; Putnam et al., 2014; Stadtler et al., 2013).

Building on this view, research using news headlines from the internet has shown that real news corrections of fake news improve memory and belief accuracy when fake news reminders precede them (Kemp et al., 2022a; Wahlheim et al., 2020). Such reminder effects were shown to reflect both improved awareness of conflicting details along with enhanced recollection of their relationship and veracity. Conditional analyses provided evidence reminders promoted associative encoding of true and false details that depended on subsequent recollection that fake news had been corrected. When participants did not recollect corrections, the familiarity of fake news led to intrusion errors and less accurate perceptions of the veracity of remembered details. Findings such as these have been taken to suggest that remindings of earlier information during new learning can effectively counteract proactive interference as long as detected changes promote subsequent recollection (for a review, see Wahlheim et al., 2021). From this perspective, fake news reminders have the potential to enhance recollection by promoting integrative encoding and increase familiarity-based errors across items within a task. Consequently, aggregate assessments of memory and belief updating depend on how often recollection-based retrieval is engaged at test.

The familiarity backfire and integrative encoding accounts of misinformation repetition effects lend naturally to competing predictions regarding age-related differences in memory and beliefs. It is well established that older adults show poorer associative memory and recollection of contextual details including source information (for a review, see Park & Festini, 2017). Although older adults recollect less well than younger adults, familiarity remains relatively invariant across age groups (Jennings & Jacoby, 1993). From the familiarity backfire perspective, the increased familiarity of repetitions should lead to more memory errors for older than younger adults because older adults would be less able to use recollection to oppose such familiarity. Indeed, research has shown the repetitions of items from non-target sources leads to worse recognition memory for target items for older adults and better recognition memory for those items for younger adults (Jacoby, 1999). The integrative encoding perspective also predicts poorer memory for older adults because associative memory for co-activated details should be impaired. However, this perspective predicts that older adults should still benefit from repetitions when they promote successful detection of corrections that leads to subsequent recollection of true and false details. Research has shown this pattern of age differences in memory updating studies using paired associate learning (Wahlheim, 2014) and everyday events (Wahlheim & Zacks, 2019).

However, to our knowledge, no studies have compared age-related differences between older and younger adults in the context of repeating misinformation. Previous work shows that older adults are more likely to make false memory and belief errors under conditions that motivate familiarity-based processes. For example, a study examining the mechanisms of belief updating showed that older adults were worse at sustaining post-correction beliefs that false claims were inaccurate relative to middle-aged adults (Swire et al., 2017a). Likewise, older adults have also been shown to misremember myths as facts after repeated retractions compared to single retractions to a greater extent than younger adults (Skurnik et al., 2005). If beliefs are partly based on memory for veracity information learned from corrections, as suggested by recent studies (Kemp et al., 2022a; Wahlheim et al., 2020), then predictions about age differences from the familiarity backfire and integrative encoding accounts for memory outcomes should extend to judgments of perceived accuracy.

The Present Study

The primary aim of the present study was to characterize age-related differences in post-correction fake news repetition effects on memory for real news details and beliefs in retrievals of those details. We accomplished this using a three-phase fake news correction paradigm in which participants 1) rated the perceived accuracy of true and fake news headlines of unclear veracity, 2) attempted to detect fake news corrections when reading real news headlines that affirmed prior real news and corrected prior fake news, and 3) completed either a recognition memory (Experiment 1) or cued recall (Experiment 2) test that assessed memory for real news details, perceived accuracy of the retrieved details, whether those details corrected fake news, and memory for the fake news details. The headlines were fake news from the internet and verified real news corrections issued by fact-checking websites. We manipulated fake news exposure by presenting headlines either once or thrice during the initial truth discernment phase. This procedure allowed us to assess post-correction fake news repetition effects on memory and the perceived accuracy of retrieved details as well as conditional retrieval of real news based on whether fake news was also retrieved. Finally, we evaluated how fake news repetitions influenced contributions of recollection and familiarity to retrieval using a hierarchical Bayesian Multinomial Processing Tree (MPT) approach.

We tested several overarching hypotheses regarding fake news repetition effects on memory and beliefs as well as potential interactions with age. Based on our prior studies of fake news correction effects on memory (Kemp et al., 2022ab; Wahlheim et al., 2020), we expected repeating fake news to improve detection of corrections and enhance subsequent memory for real news to the extent that participants could recollect that fake news had been corrected. In the absence of such recollection, we expected fake news repetitions to increase proactive interference, and thus more intrusion errors, consistent with our findings in episodic memory updating using paired-associated learning tasks (Wahlheim, 2014; Wahlheim et al., 2019). Furthermore, prior work shows that repeating items from a non-target source improves recognition for younger adults and hurts recognition for older adults because of a selective age-related recollection deficit (Jacoby, 1999). Based on these findings, we expected older adults to benefit less from the potential for fake news repetitions to promote detection of corrections and subsequent recollection of the relationship between real and fake news details. We expected age-related recollection differences to be captured by MPT estimates.

Finally, based on our work showing higher belief accuracy when veracity information is better recollected, we expected higher perceived accuracy ratings for correct than incorrect real news retrievals and for ratings for correct recall to be higher when participants could recollect that fake news was earlier corrected. Although we expected this pattern for both age groups, it is possible that older adults could show lower belief accuracy if their recollection deficit undermines their use of memory for corrections as a basis for their perceptions of headline detail veracity. We also explored changes in perceived accuracy ratings from the initial to post-correction ratings to determine if such changes better reflected the provision of veracity information during the correction phase for correct retrievals of new details. If so, this would suggest that participants based their beliefs strongly on memory for corrections.

Experiment 1

Experiment 1 characterized the effects of repeated fake news exposure and real news corrections on memory and perceived accuracy of real and fake news headlines in younger and older adults. We manipulated the fake news exposure frequency before corrections then tested younger and older adults' recognition and perceived accuracy of retrieved details. This allowed us to assess the contributions of recollection and familiarity-based retrieval and to test predictions from leading theories of misinformation correction effects described above.

Methods

All stimuli, data, and analysis scripts are available here: https://osf.io/vqwtu/. These experiments were approved by the Institutional Review Board at The University of North Carolina at Greensboro (UNCG; IRB-FY21-179).

Participants

The stopping rule was to obtain useable data from at least 102 younger and 102 older adults. This sample size also allowed us to administer each of three experimental formats equally across subjects. We recruited participants online from Prolific (www.prolific.ac). We used the pre-screening settings to ensure that participants had a high approval rating, were roughly equal with gender, indicated the United States as their nationality, resided in the United States at the time of testing, and whose ages were between 18-35 years (younger adults) or 65-75 years (older adults). Participants received $10 for completing two sessions.

The final sample included 102 younger adults (62 women, 40 men) ages 18-29 years (*M* = 22.70, *SD* = 2.92) and 102 older adults (62 women, 40 men) ages 65-75 years (*M* = 69.10, *SD* = 2.76). We excluded data from 33 younger adults and 22 older adults. Of those participants, 24 younger and 11 older did not return for the second session, four younger and four older were exposed to the procedure in the first session multiple times because they re-opened the first session before starting the second session, four younger and three older did not complete the first session, one younger and three older completed the study after we had reached our target sample size, and one older did not complete the second session.

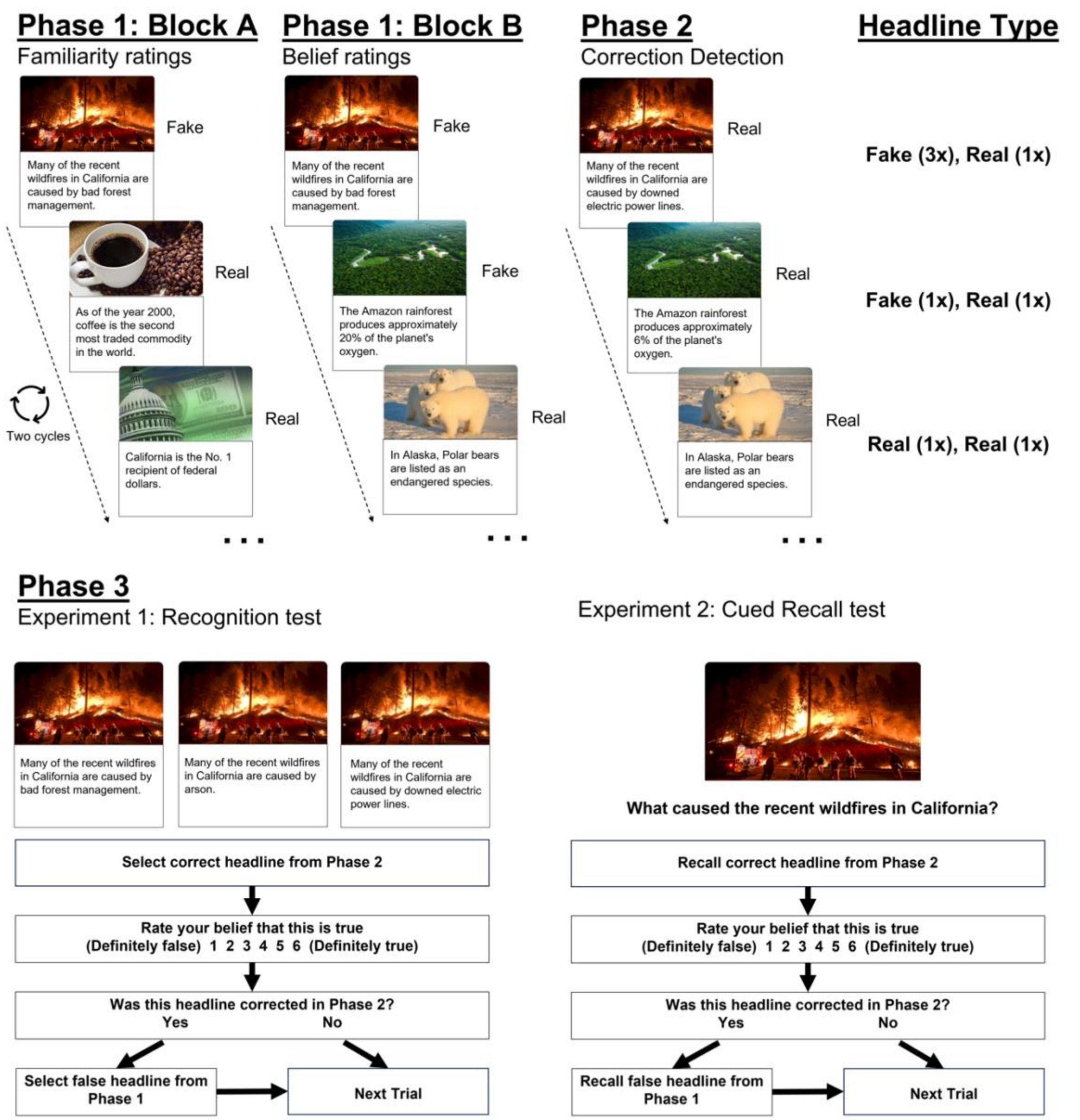
Design

This experiment used a mixed factorial design. It included Age as a between-subjects variable with younger and older adults as levels. It also included Headline Type as a within-subjects variable with three levels varying based on the relationship between headline veracity in Phases 1 and 2. First, a repeated real news condition included one real news headline in Phase 1 and a repetition of that headline in Phase 2 [Real (1×), Real (1×)]. In addition, a single-exposure fake news correction condition included one fake news headline in Phase 1 and a real news correction of that headline in Phase 2 [Fake (1×), Real (1×)]. Finally, a repeated-exposure fake news correction condition included three presentations of the same fake news headline in Phase 1 and a real news correction of that headline in Phase 2 [Fake (3×), Real (1×)].

Materials and Procedure

Figure 10 displays the example stimuli, experimental conditions, and procedures from Experiment 1. The stimuli set included 60 headline pairs from the fact-checking websites such as politifact.com and snopes.com. Each set included a real and fake news headline on the same unique topic. Fake news headlines included a false detail, and real news headlines included a true detail that corrected the false detail. All fake news headlines were originally portrayed by the media as being true. The headline format resembled news updates on internet search engines. Real and fake news headlines about a topic appeared below an image related to the topic. Of the 60 statement pairs, 45 served as critical items and 15 served as fillers that appeared in the first phase (see below). We counterbalanced critical items by rotating three sets of 15 pairs through conditions, creating three experimental formats. Headlines appeared equally often in each condition across participants.

Figure 10. Schematic of the Procedure



Note. A schematic overview of the headline types and trial structures in two experiments. In Phase 1, participants were exposed to real news and fake news headlines that were presented once or thrice. To do this, Phase 1 was divided into two seamless blocks. In Block A, participants were shown two presentations of fake news headlines that would again be shown in Block B for its final presentation (x3) and real news filler headlines and rated them for familiarity. In Block B, participants were shown fake news headlines that had been presented before, new fake news headlines, and new real news headlines and rated them for perceived truthfulness. During Phase 2, participants were shown headlines that re-affirmed real news or corrected fake news and indicated when fake news was corrected. A difference between experiments was the trial structure in Phase 3. In Experiment 1, participants were given a 3AFC task and were first asked to identify Phase 2 real news details, rate the perceived truthfulness of the headline details they recognized, then indicated those for which fake news was corrected in Phase 2, and for those, attempted to recognize the Phase 1 fake news. The red border indicates the participants selection. In Experiment 2, participants first recalled Phase 2 real news details, rated the truthfulness of the headline details they reported, indicated those for which fake news was corrected in Phase 2, and for those, attempted to recall the Phase 1 fake news.

We controlled stimulus presentation using Inquisit software (*Inquisit 5*, 2016). Participants completed the experiment on their own laptops or desktop computers unsupervised. The experiment included three phases. In each phase, stimuli appeared in a fixed random order with the restriction that no more than three headlines from the same condition appeared consecutively. The average list position for each condition was equated to control for serial position effects. During interstimulus intervals, a blank screen appeared for .5 s, then a button labeled “Next” appeared in the center of the screen. To ensure task engagement, participants were required to click that button with a mouse to advance to the next trial.

To manipulate fake news exposure in Phase 1, we created two seamless blocks and distributed repetitions of fake news headlines across those blocks. Before the first block, an instruction screen told participants that real and fake news headlines would appear and that they should study the headlines for a later test. We presented an equal number of real and fake news headlines in the first block by intermixing real news filler items with the first two presentations of fake news headlines from the condition with three fake news presentations. The first block included 30 unique headlines (15 fake news critical items and 15 real news fillers) that appeared once in each of two cycles. All 30 headlines appeared once in the first cycle before any headline repeated in the second cycle (60 total presentations in Phase 1). Participants indicated their familiarity with every headline from 1 (Definitely Unfamiliar) to 6 (Definitely Familiar) by clicking on response boxes displayed on the screen.

Before the second block in Phase 1, another instruction screen told participants that real and fake news headlines would appear and that they should study the headlines for a later test. They were also informed that some upcoming statements would be repetitions from the first block whereas other statements would be new. The second block included 60 headlines with an equal number of real and fake news types. Thirty real news headlines comprised 15 critical items that would be repeated in Phase 2 (Repeated Real News) and 15 repetitions of fillers from the first block. Thirty fake news headlines corresponded with the two fake news exposure conditions: One set included 15 new fake news headlines that only appeared once in the second block and would be corrected in Phase 2 (Corrected Fake News [1×]). The other set included the third presentation of the 15 fake news headlines that appeared twice before in the first block and would also be corrected in Phase 2 (Corrected Fake News [3×]). In the second block, participants rated the truthfulness of headlines on a scale from 1 (Definitely False) to 6 (Definitely True) by clicking on response boxes displayed on the screen. Each headline appeared for 8 s.

Before Phase 2, an instruction screen told participants that they would read real news headlines that would either repeat real news or correct fake news from Phase 1. It further instructed participants to indicate whether headlines corrected fake news. Phase 2 included 45 headlines, with 15 headlines corresponding to each of the Headline Type conditions. To indicate which headlines corrected fake news, participants responded "Yes" or "No" by clicking responses boxes displayed on the screen at which point the options disappeared. Each headline appeared for 8 s.

After Phase 2, an instruction screen told participants to start the second session after 48 hours and no later than 73 hours. However, our method of deploying the sessions was limited by features of the online platform. Consequently, some participants started slightly earlier than 48 hours after completing the first session. This occurred when participants started the second session two days later, as instructed, but earlier than when they started the first session. The average number of hours between sessions was not significantly different between younger adults (M = 51.81, SD = 6.23, *Range* = [43.16 – 71.37]) and older adults (M = 52.11, SD = 6.66, *Range* = [43.27 – 72.67]), $t(202) = 0.34$, $p = .73$. During this interval, a research assistant checked each data file to ensure that the participants completed all the trials in the first session. Upon verifying this, participants were then manually placed on a custom list that granted them access to the second session. Participants were notified twice through the Prolific messaging portal about returning to complete the second session, once the morning before the session became available, once immediately after the session became available, and again the next morning for participants who had not started the study by then.

Before Phase 3, an instruction screen told participants that their task would be to answer questions about their memory for and perceived accuracy of the headlines they read in the first session. Phase 3 included a three alternative forced choice (3AFC) recognition memory test of the 45 critical headline topics. On each trial, three headlines appeared, each below the associated image from the first session. The headlines included: real news from Phase 2, fake news that appeared in Phase 1 for the correction conditions (which was new for the repeated real news condition, because fake news headlines never appeared in Phase 1 for this condition), and fake news including a detail that we generated but that was not presented during either phase. Including the second fake news headline with an extra-experimental detail allowed us to precisely examine source confusion between fake news and real news (see Statistical Methods). The position of the real news headline was counterbalanced so that it appeared equally often in each position across trials and never appeared in the same position more than twice sequentially.

On each trial, participants first attempted to select the real news headline from Phase 2. Next, they rated the truthfulness of the chosen headline using the Phase 1 scale: 1 (Definitely False) to 6 (Definitely True) by clicking response boxes on the screen. Following this, participants indicated via key press if real news in Phase 2 had corrected fake news from Phase 1 by responding “Yes” (1) or “No” (0). After responding “yes,” they attempted to indicate which of the remaining headlines was from Phase 1. After responding “no,” they advanced to the next trial.

Statistical Methods

In both experiments, we performed all statistical tests using R software (R Core Team, 2022). To examine the effects of varying headline types, we fitted linear and logistic mixed-effects models using functions from the lme4 package (Bates et al., 2015). Based on signal detection theory (SDT; Green & Swets, 1966) we also characterized participants' detection of corrections during Phase 2 and subsequent memory that corrections had been detected during Phase 3 in terms of discrimination (d') and response bias (c). We calculated these parameter estimates by computing hit and false alarm rates for each participant and using the dprime function from the psycho package (Makowski, 2021) to estimate the parameters. We performed Wald's $\chi 2$ hypothesis tests using the Anova function of the car package (Fox & Weisberg, 2019) and post hoc comparisons using the Tukey method in the emmeans package (Lenth, 2021), which controlled for multiple comparisons.

We additionally fit hierarchical Bayesian MPT models using the TreeBUGS package (Heck et al., 2018) to estimate the contributions of recollection and familiarity to responses on the first step of the Phase 3 test procedure in the corrected fake news conditions. The models estimate the probability of these latent cognitive parameters based on the frequency of each response type (i.e., correct recognition/recall of real news from Phase 2, incorrect recognition/intrusions of fake news from Phase 1, and incorrect recognition/recall of details that never appeared). Following similar work (Bartsch et al., 2019; Kemp et al., 2022a) based on dual process models of memory (Jacoby, 1999), we assumed that participants could correctly recognize/recall real news headlines based on recollection (*Pr*). When recollection fails (1 – *Pr*), participants may be familiar with real and fake news headline details from earlier phases (*Pf*), leading to equal probabilities of guessing the details from both headline types. Finally, without familiarity (*Pf*), participants guess with equal probability among the three response types. The parameters of interest are thus *Pr* and *Pf*, whereas the two guessing parameters were fixed to 0.5 to achieve model identifiability.

The MPT models are hierarchical because they estimate parameters for each participant and are Bayesian because they estimate the parameters' posterior distributions based on uninformative priors and the data using Markov Chain Monte Carlo sampling. Each model was conducted with 4 chains of 100,000 iterations, with 20,000 iterations for adaptation, 2,000 iterations for burn-in, and a thinning factor of 5. The results showed model convergence and adequate fit to the data. This enabled comparison of the posterior distributions to determine if differences in the parameter estimates across conditions were credible (i.e., the 95% credibility intervals, CIs, of the differences do not overlap with 0).

All models included Age and Headline Type as fixed effects as well as Participants and Items as random intercept effects. Given the self-paced access to Phase 3, we controlled for study-test delay in the mixed-effects models of Phase 3 recall performance by including retention time as a fixed effect. We did not include this variable when doing so prevented the models from converging. The complete model specifications are available in the analysis scripts on the OSF. The significance level was $\alpha = .05$.

Results and Discussion

Perceived Accuracy: Phase 1

We compared younger and older adults' initial perceived accuracy ratings for real and fake news headlines before corrections (Table 7). A model with Age and Headline Type and as fixed effects indicated significant effects of Age, $\chi^2(1) = 48.72$, $p < .001$, and Headline Type, $\chi^2(2) = 82.39$, $p < .001$, and no interaction $\chi^2(2) = 4.80$, $p = .09$. Younger adults made higher overall ratings than older adults. Both groups made higher ratings for real than fake news, smallest *z* ratio = 4.77, $p < .001$, and for fake news appearing thrice compared to once, *z* ratio = 4.31, $p < .001$. These results suggested that younger adults were less skeptical in their beliefs, both age groups generally discerned true from false details, and that repeating fake news led to illusory perceptions of truth (Hasher et al., 1977; Hassan & Barber, 2021).

Table 7. Baseline Perceived Accuracy of Real and Fake News

		News Headline Type		
Experiment	Age	Real 1×	Fake 1×	Fake 3×
1	Younger	3.81 [3.64, 3.97]	3.54 [3.37, 3.71]	3.72 [3.55, 3.89]
	Older	3.47 [3.30, 3.63]	3.11 [2.94, 3.28]	3.23 [3.06, 3.39]
2	Younger	3.69 [3.52, 3.85]	3.39 [3.23, 3.56]	3.51 [3.34, 3.67]
	Older	3.50 [3.34, 3.66]	3.23 [3.07, 3.40]	3.31 [3.14, 3.47]

Note. Bootstrap 95% confidence intervals are displayed in brackets.

Correction Classifications: Phases 2 and 3

Phase 2 (Detecting Corrections)

We assessed younger and older adults' detection of fake news corrections in Phase 2 and subsequent recognition of fake news and recollection that it was corrected in Phase 3 to characterize the extent to which these processes played a role in real news recognition. We computed probabilities of "yes" responses separately for correction classifications in each phase across headline type conditions (Table 8). Note that these responses are correct classifications for both fake news headline types and false alarms for real new headlines. Considering both of these response types allowed us to determine how well participants could discriminate between headline types. To assess such discrimination and the extent to which classifications reflected response biases, we estimated d' and c parameters, respectively, using complementary signal detection analyses (Figure 11). We calculated parameter estimates by treating "yes" responses to each corrected fake news headline type as separate hit rates.

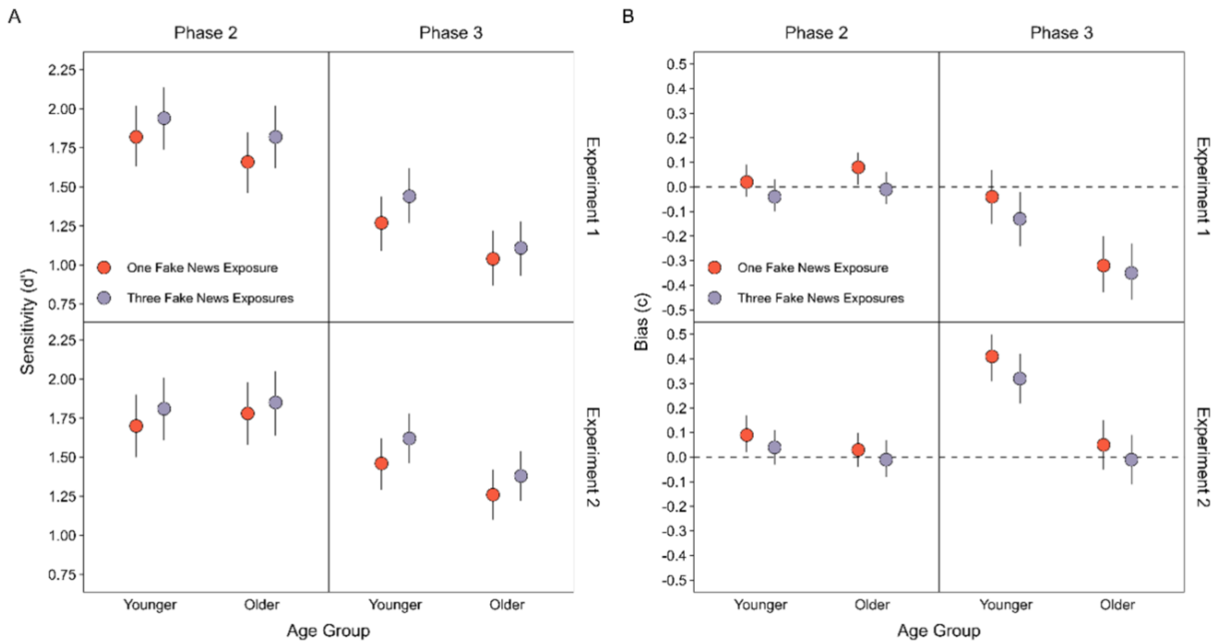
Table 8. Correction Classifications for Real and Fake News

			Headline Type		
Experiment	Phase	Age	Real News	Fake News 1×	Fake News 3×
1	2	Younger	.17 [.14, .21]	.83 [.79, .86]	.85 [.81, .88]
		Older	.18 [.15, .21]	.78 [.74, .82]	.83 [.80, .86]
	3	Younger	.26 [.22, .31]	.77 [.72, .81]	.82 [.78, .85]
		Older	.40 [.34, .46]	.82 [.77, .85]	.84 [.80, .87]
2	2	Younger	.18 [.15, .21]	.79 [.76, .83]	.81 [.77, .84]
		Older	.18 [.15, .22]	.83 [.79, .86]	.84 [.81, .87]
	3	Younger	.09 [.07, .12]	.64 [.58, .70]	.69 [.63, .75]
		Older	.23 [.18, .28]	.74 [.68, .79]	.78 [.73, .82]

Note. Bootstrap 95% confidence intervals are displayed in brackets.

We compared "yes" responses (correction classifications) using models with Age and Headline Type as fixed effects for each phase (Table 8). The model for detections in Phase 2 indicated no significant effect of Age, $\chi^2(1) = 1.57$, $p = .21$, and a significant effect of Headline Type, $\chi^2(2) = 2435.34$, $p < .001$, and a significant interaction, $\chi^2(2) = 6.11$, $p = .047$. Both age groups discriminated fake news corrections from real news affirmations as shown by significantly higher probabilities for both fake news correction types than for real news repetitions, smallest *z* ratio = 29.82, $p < .001$. The fake news repetition effect was larger for older adults, as correct classification was significantly higher when fake news appeared thrice than once for older, *z* ratio = 3.39, $p = .002$, but not younger, *z* ratio = 1.52, $p = .28$, adults.

Figure 11. Signal Detection Parameter Estimates for Correction Classifications in Phases 2 and 3



Note. Probabilities of signal detection estimates for correction classifications in phases 2 and 3 for each correction headline type condition. Points are probabilities estimated from mixed effects models; error bars are 95% confidence intervals.

We further examined this age difference by comparing signal detection parameter estimates between fake news correction conditions. A model including Age and Headline Type as fixed effects for d' (Figure 11A, top left) indicated no significant effect of Age, $\chi^2(1) = 1.06$, $p = .30$, a significant effect of Headline Type, $\chi^2(1) = 16.49$, $p < .001$, and no significant interaction, $\chi^2(1) = 0.61$, $p = .43$, showing that discriminability was higher for corrections of fake news that appeared thrice compared to once. In addition, a comparable model for c indicated no significant effect of Age, $\chi^2(1) = 0.87$, $p = .35$, a significant effect of Headline Type, $\chi^2(1) = 16.55$, $p < .001$, and no significant interaction, $\chi^2(1) = 0.62$, $p = .43$, showing that participants adopted a more liberal response biases for corrections of fake news that appeared thrice compared to once. These results show that when accounting for false alarms to real news repetitions, both age groups showed comparable detection of corrections.

Phase 3 (Remembering Corrections)

Turning to correction classifications in Phase 3 (Table 8), the model indicated significant effects of Age, $\chi^2(1) = 6.90$, $p < .01$, and Headline Type, $\chi^2(2) = 1486.73$, $p < .001$, that were qualified by a significant interaction, $\chi^2(2) = 14.49$, $p < .001$. As in Phase 2, younger and older adults discriminated topics that included corrected fake news from those that included affirmed real news, as response probabilities were higher for both fake news headlines than real news headlines, smallest z ratio = 20.86, $p < .001$. Unlike Phase 2, younger adults showed higher probabilities for fake news that appeared thrice than once, z ratio = 3.26, $p < .01$, whereas older adults showed no difference, z ratio = 1.90, $p = .14$. Together, these results indicate better overall discrimination for younger than older adults, especially when fake news appeared thrice before corrections.

Signal detection parameter estimates compared using the same modeling approach as in Phase 2 verified this discrimination difference. The model for d' (Figure 11A, top right panel) indicated significant effects of Age, $\chi^2(1) = 5.37$, $p = .02$, and Headline Type, $\chi^2(1) = 12.06$, $p < .001$, and no significant interaction, $\chi^2(1) = 2.71$, $p = .10$, showing that younger adults identified topics associated with fake news corrections better than older adults, and such identification by both groups was better when fake news appeared thrice compared to once. In addition, a comparable model for c indicated significant effects of Age, $\chi^2(1) = 9.90$, $p < .01$, and Headline

Type, $\chi^2(1) = 12.05$, $p < .001$, and no significant interaction, $\chi^2(1) = 2.73$, $p = .10$, showing that older adults adopted a more liberal response bias than younger adults and that both groups adopted more liberal response biases for corrections of fake news that appeared thrice compared to once. Together, these results show that when accounting for false alarms to repeated real news headline topics, younger adults showed uniformly better discrimination and more conservative responding than older adults when identifying topics for which fake news was corrected.

Overall 3AFC Recognition Memory: Phase 3

We examined the proactive effects of fake news exposure prior to corrections on subsequent memory accuracy by examining recognition memory for real and fake news details in Phase 3. We assessed memory accuracy by comparing correct recognition of real news headlines as well as false and correct recognition of fake news headlines using separate models for each memory measure with Age and Headline Type as fixed effects.

Correct Recognition of Real News

Figure 12A displays correct recognition of real news, which refers to when participants chose the real news headline from the three alternatives. The model indicated no significant effect of Age, $\chi^2(1) = 0.95$, $p = .33$, a significant effect of Headline Type, $\chi^2(2) = 77.80$, $p < .001$, and no significant interaction, $\chi^2(2) = 0.84$, $p = .66$. Repeating real news led to higher accuracy than presenting fake news before corrections, smallest z ratio = 7.37, $p < .001$. There were no accuracy differences between fake news that appeared thrice or once, z ratio = 20.86, $p < .001$. Unsurprisingly, these results show that repeating real news improved memory accuracy compared to when fake news created response competition for real news details. More interesting was that overall accuracy did not vary with the number of fake news presentations nor age. Based on earlier studies indicating roles for detecting and recollecting corrections in subsequent memory, the latter result suggests that repeating fake news created offsetting facilitation and interference effects. We summarize those effects below.

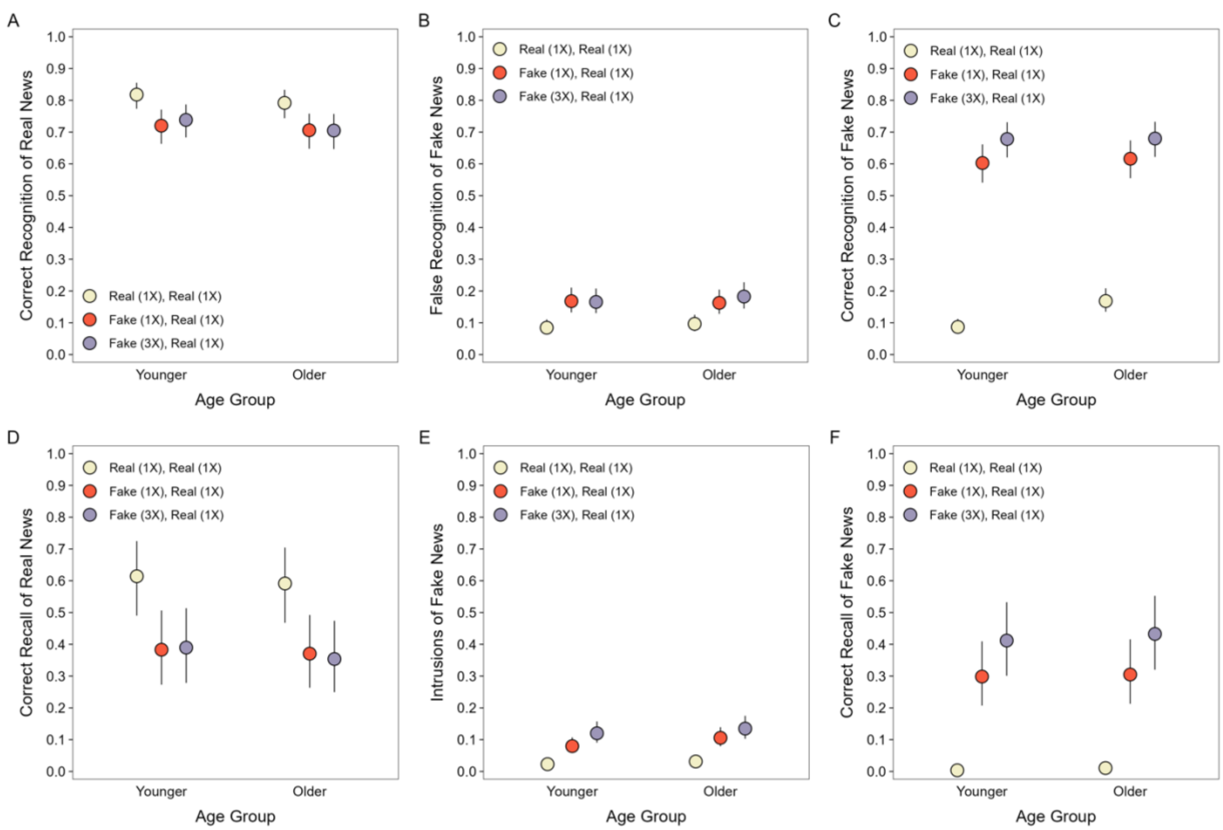
False Recognition of Fake News

Figure 12B displays false recognition of fake news, which refers to when participants chose the fake news headline from the three alternatives. In these comparisons, the repeated real news condition served as a baseline index of falsely recognizing fake news details that never appeared in the experiment. In contrast, false recognition of details that originated in Phase 1 reflect memory errors in which veracity information was not retrieved. The model indicated no significant effect of Age, $\chi^2(1) = 0.26$, $p = .61$, a significant effect of Headline Type, $\chi^2(2) = 117.85$, $p < .001$, and no significant interaction, $\chi^2(2) = 2.01$, $p = .37$. False recognition of headlines that appeared in Phase 1 was significantly above baseline, smallest *z* ratio = 9.14, $p < .001$, and did not differ based on exposure frequency, *z* ratio = 0.91, $p = .64$. As for correct recognition, the lack of a fake news repetition effect on false recognition points toward offsetting effects that depended on detecting and recollecting corrections.

Correct Recognition of Fake News

Figure 12C displays correct recognition of fake news, which refers to when participants indicated that fake news had been corrected and then chose the fake news headline from Phase 1 from the two alternatives remaining after the classification. In these comparisons, correctly recognizing the fake news in the repeated real news condition reflects a semantic memory error, whereby participants spontaneously recognized fake new details without having seen them earlier in the experiment. The model indicated no significant effect of Age, $\chi^2(1) = 2.07$, $p = .15$, a significant effect of Headline Type, $\chi^2(2) = 1607.28$, $p < .001$, and a significant interaction, $\chi^2(2) = 33.98$, $p < .001$. Both age groups recognized fake news as such significantly more often when it appeared thrice than once, smallest *z* ratio = 3.35, $p < .01$, and when it was corrected compared to when it did not appear in the experiment, smallest *z* ratio = 22.88, $p < .001$. The interaction reflected older adults being more likely than younger adults to recognize fake news that did not appear in the experiment, *z* ratio = 4.46, $p < .001$.

Figure 12. Overall Response Probabilities Indicating Memory for Real and Fake News Headlines in Experiment 1 (Upper Panels) and Experiment 2 (Lower Panels)



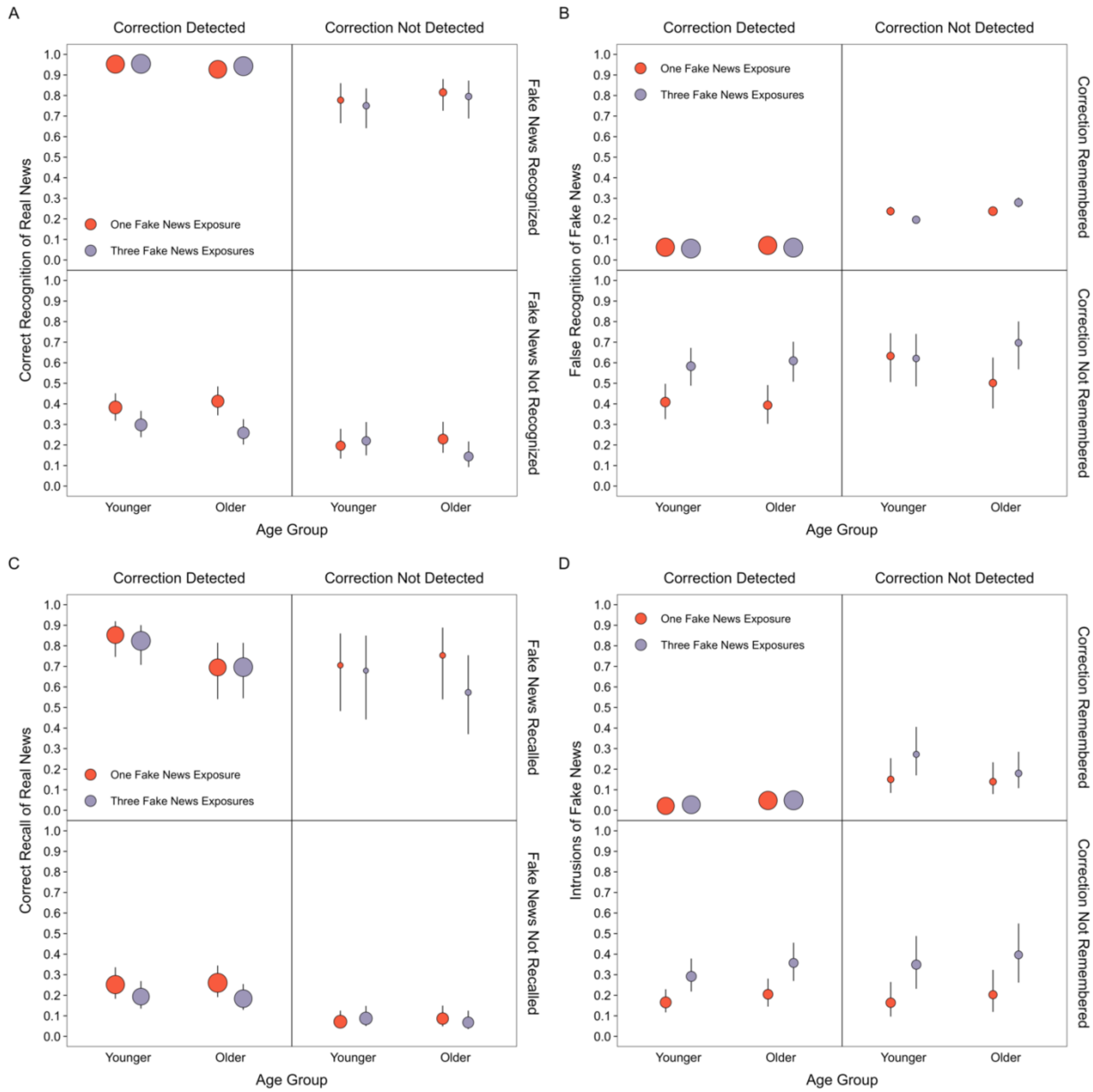
Note. Probabilities of real news correct recognition and recall, fake news false recognition and intrusions, and fake news correct recognition and recall in Phase 3 for each headline type condition. Points are probabilities estimated from mixed effects models; error bars are 95% confidence intervals.

Recognition in Phase 3 Conditionalized on Correction Classifications from Phases 2 and 3

Correct Recognition of Real News

To characterize the offsetting effects of fake news repetitions that depended on detection and recollection of corrections, we conducted conditional analyses of correct and false recognition (Figure 13). We first examined the extent to which correct recognition of real news headlines depended on participants detecting corrections in Phase 2 and recognizing the Phase 1 fake news headlines in Phase 3 by conditionalizing real news recognition for the two correction headline types on those variables (Figure 13A). The observation frequencies corresponding to different point sizes appear in Appendix C in Table S5 (top rows). We used a model including Age, Correction Detection (Phase 2), Fake News Recognition (Phase 3), and Headline Type as fixed effects. Note that fake news recognition had two levels, correct recognition hits (fake news recognized) and incorrect recognition misses (fake news not recognized). To simplify exposition, the complete model results appear in Table S6. Below, we highlight key main effects and describe their qualifications by other variables by reporting results from two-way models for each combination of correction detection and fake news recognition including fixed effects of Age and Headline Types. We repeat this approach for analyses of false recognition of fake news and the comparable conditional analyses of recall of real news and intrusions of fake news in Experiment 2.

Figure 13. Conditional Response Probabilities Indicating Memory for Real and Fake News Headlines Based on Correction Classifications in Phases 2 and 3 for Experiment 1 (Panels A and B) and Experiment 2 (Panels C and D)



Note. Probabilities of real news correct recognition and recall and fake news false recognition and intrusions in Phase 3 conditionalized on correction classifications for each correction headline type condition. Points are probabilities estimated from mixed effects models;

error bars are 95% confidence intervals. Point sizes indicate for each cell the proportion of observations, which are also displayed in Supplementary Tables S5 and S9.

Consistent with prior findings, detecting corrections and subsequently recognizing fake news headlines were both associated with better recognition of real news headlines, smallest $\chi^2(1) = 77.65$, $p < .001$. These effects were qualified by interactions including various combinations of all fixed effects, smallest $\chi^2(1) = 4.50$, $p = .03$. The differences associated with detecting corrections and recognizing fake news can be seen in Figure 13A by looking between columns and rows, respectively. The benefit to real news recognition was substantially larger when fake news was recognized than when corrections were detected, and real news recognition was best when both occurred. When corrections were detected and fake headlines were recognized as such (top left panel), the real news recognition benefit was greater for younger than older adults, $\chi^2(1) = 4.16$, $p = .04$, and did not differ based on fake news exposures, $\chi^2(1) = 0.74$, $p = .39$. In contrast, when corrections were detected but fake headlines were not recognized as such (bottom left panel), real news recognition was significantly lower when fake news had appeared thrice compared to once, $\chi^2(1) = 19.43$, $p < .001$, indicating that additional fake news exposures created more proactive interference. No other significant effects were observed from these models nor from the remaining models (both right panels), largest $\chi^2(1) = 3.71$, $p = .05$. These results suggest that additional fake news exposure did impair overall real news recognition, but this impairment was offset by the benefits associated with better detection and recollection of corrections.

False Recognition of Fake News

We next examined the extent to which false recognition of fake news headlines depended on participants detecting corrections (Phase 2) and remembering corrections (Phase 3) using the same modeling approach as in the previous analyses (Figure 13B). The two levels of remembering corrections were correct and incorrect Phase 3 classifications of headline topics being corrected in Phase 2 (Correction Remembered and Correction Not Remembered, respectively). We could not conditionalize these responses on fake news recognition because there were no response options after being incorrectly identified as real news. The observation frequencies corresponding to point sizes appear in the bottom rows of Table S5, and the complete model results appear in Table S7.

Detecting and subsequently remembering corrections were both separately associated with fewer false recognitions of fake headlines, smallest $\chi^2(1) = 113.40$, $p < .001$. There was also a significant effect of Headline Type, $\chi^2(1) = 7.74$, $p < .01$, and two-way interactions including different combinations of all variables except Age, smallest $\chi^2(1) = 20.05$, $p < .001$. The differences associated with detecting and remembering corrections can be seen in Figure 13B by looking between columns and rows, respectively. When corrections were detected and subsequently remembered (top left panel), the reductions in false recognition did not differ between age groups or exposures to fake news in Phase 1; no effects were significant, largest $\chi^2(1) = 1.47$, $p = .23$. In contrast, when corrections were detected but not subsequently remembered (bottom left panel), false recognition was significantly higher when fake news appeared thrice compared to once, $\chi^2(1) = 24.17$, $p < .001$, reflecting greater proactive interference following more fake news exposures. No other significant effects were observed from this model, largest $\chi^2(1) = 0.30$, $p = .59$, nor from the model for undetected corrections that were remembered, largest $\chi^2(1) = 1.48$, $p = .22$ (top right panel). But when undetected corrections were also not remembered (bottom right panel), a significant interaction, $\chi^2(1) = 4.47$, $p = .03$, indicated significantly higher fake news false recognition following three than one fake news exposures for older adults, *z* ratio = 2.84, $p < .01$, but not younger adults, *z* ratio = 0.17, $p = .86$.

Recollection and Familiarity Process Estimates for Phase 3 Recognition

Table 9 shows the parameter estimates for recollection and familiarity from the MPT model. For recollection, there were no credible effects of age or repetition. For familiarity, older adults showed credibly lower familiarity than younger adults for fake news presented once (0.10 [0.01, 0.20]) that credibly increased when fake news was repeated thrice (0.10 [0.01, 0.19)]. This age difference suggests that older adults exhibited increased familiarity with fake news following repeated exposure that contributed to their false recognition of fake news.

Table 9. Posterior Parameters and Differences Estimated from MPT Models

			Headline Type		
Experiment	Parameter	Age	Fake News 1×	Fake News 3×	*Repetition effect*
1	Recollection	Younger	0.47 [0.39, 0.55]	0.49 [0.39, 0.58]	0.02 [-0.07, 0.10]
		Older	0.46 [0.36, 0.55]	0.44 [0.36, 0.52]	-0.02 [-0.10, 0.07]
		Age effect	0.01 [-0.11, 0.13]	0.04 [-0.08, 0.17]	
	Familiarity	Younger	0.70 [0.64, 0.76]	0.73 [0.67, 0.79]	0.03 [-0.06, 0.11]
		Older	0.61 [0.53, 0.67]	0.70 [0.64, 0.75]	**0.10 [0.01, 0.19]**
		Age effect	**0.10 [0.01, 0.20]**	0.03 [-0.05, 0.11]	
2	Recollection	Younger	0.31 [0.25, 0.37]	0.26 [0.19, 0.33]	-0.05 [-0.12, 0.02]

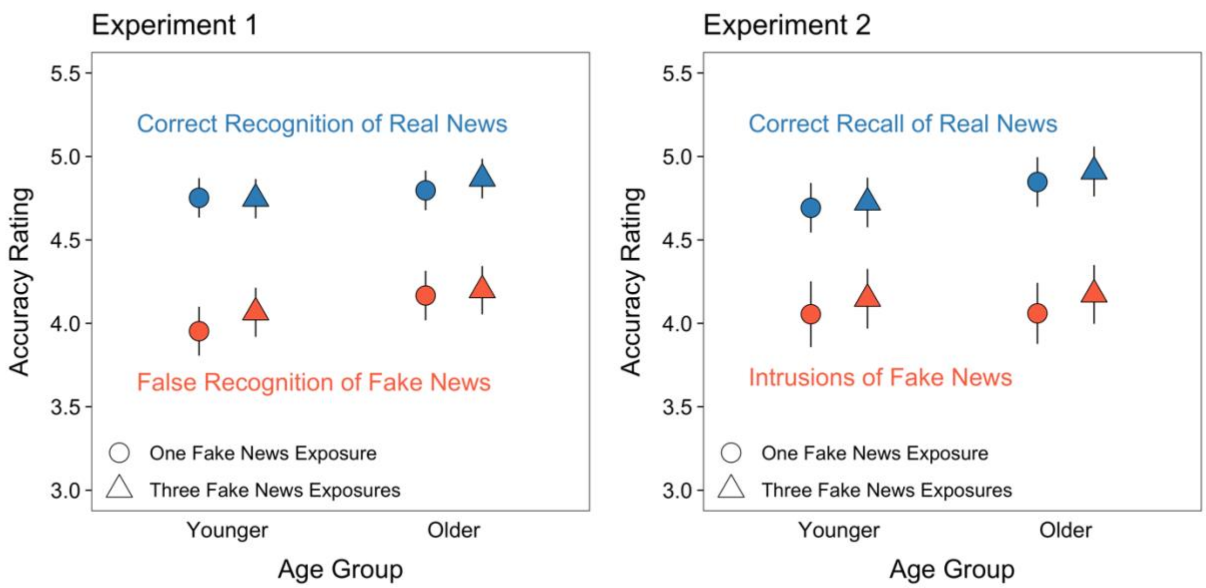
	Older	0.27 [0.21, 0.32]	0.22 [0.16, 0.28]	-0.05 [-0.11, 0.02]
	Age effect	0.04 [-0.04, 0.12]	0.04 [-0.05, 0.13]	
Familiarity	Younger	0.02 [0.00, 0.07]	0.16 [0.07, 0.23]	**0.14 [0.05, 0.22]**
	Older	0.06 [0.01, 0.14]	0.09 [0.03, 0.16]	0.03 [-0.07, 0.12]
	Age effect	-0.04 [0.12, 0.03]	0.07 [-0.04, 0.17]	

Note. Effects refer to differences between age groups and fake news repetitions. Effects in boldface are credible (i.e., their 95% confidence intervals do not overlap with 0). Bootstrap 95% confidence intervals are displayed in brackets.

Perceived Accuracy Ratings (Beliefs) for Recognized Headlines in Phase 3

We next examined differences in perceived accuracy for correction headline types that participants recognized as real news in Phase 3. This allowed us to examine if fake news exposure impacted the accuracy of beliefs. To assess belief accuracy for headline details, we compared accuracy ratings for correct recognitions of real news from Phase 2 and false recognition of fake news from Phase 1 (Figure 14, left panel). Belief accuracy was indicated by the degree to which participants perceived correctly recognized real news as more accurate than falsely recognized fake news; larger differences in accuracy ratings indicated more accurate beliefs. A model with Age, Response Type, and Headline Type as fixed effects indicated a significant effect of Response Type, $\chi^2(1) = 462.75$, $p < .001$, and no other significant effects, largest $\chi^2(1) = 2.85$, $p = .09$. The complete results appear in Table S8. Participants perceived recognized real news as more accurate than falsely recognized fake news to the same degree regardless of the amount of initial fake news exposure.

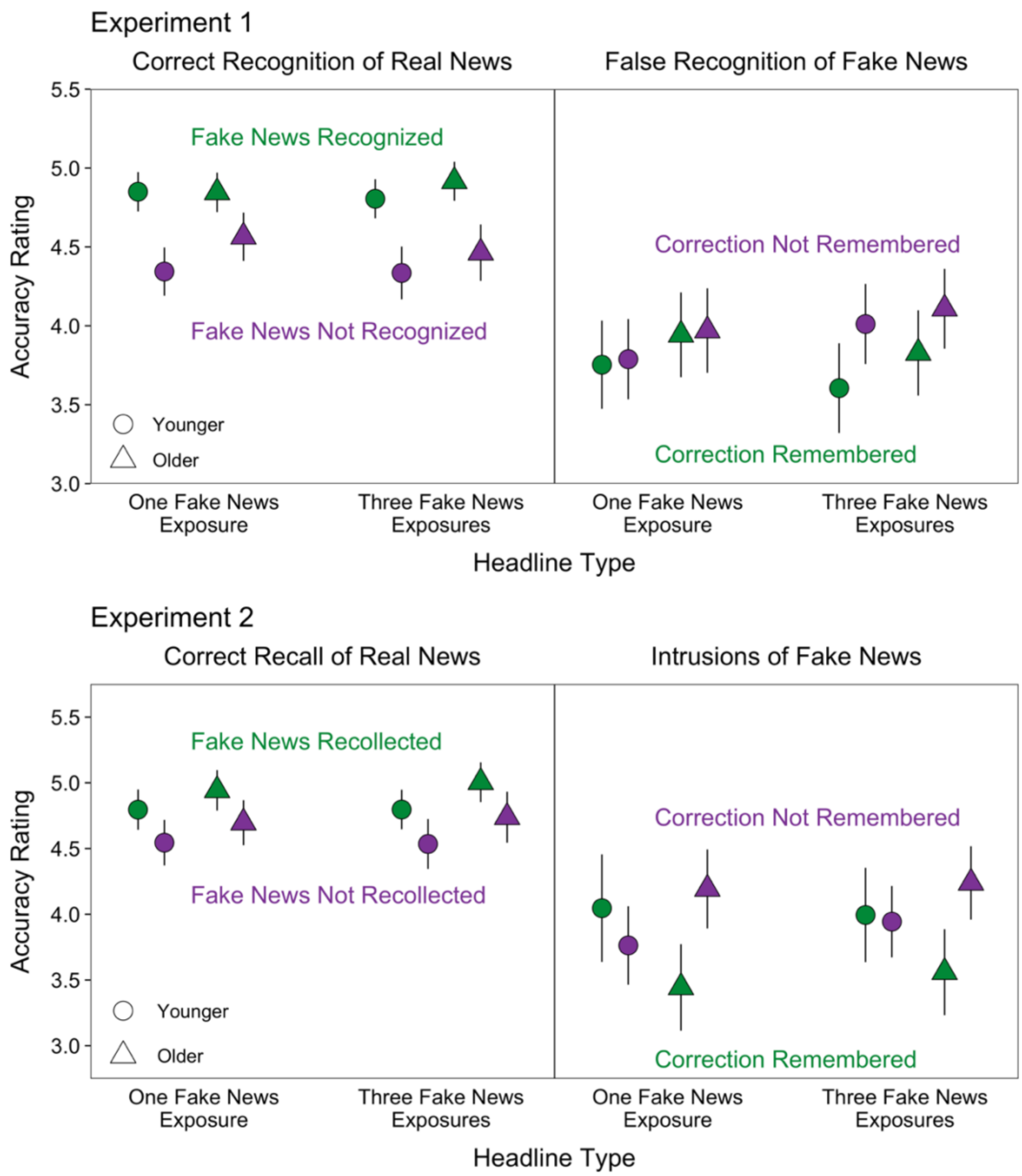
Figure 14. Perceived Accuracy Ratings for Real and Fake News Headlines for Each Correction Headline Type Condition in Phase 3: Experiments 1 and 2



Note. Probabilities of perceived accuracy ratings for real news correct recognition and recall and fake news false recognition and intrusions in Phase 3. Points are probabilities estimated from mixed effects models; error bars are 95% confidence intervals.

We further assessed the relationship between memory and beliefs by modelling accuracy ratings separately for each response type conditionalized on memory for corrections in Phase 3 (Figure 15A). The models included Age, Headline Type, and Correction Classification as fixed effects. Note that as for similar conditional analyses above, Correction Classification refers to whether fake news was recognized after corrections were identified as such for correct recognition of real news and to whether corrections were classified as such following false recognition of fake news. The complete results appear in Table S9.

Figure 15. Perceived Accuracy Ratings for Real and Fake News Headlines Conditionalized on Correction Classification for Each Correction Headline Type Condition: Experiments 1 and 2



Note. Probabilities of perceived accuracy ratings for real news correct recognition and recall and fake news false recognition and intrusions in Phase 3 conditionalized on correction classifications for each correction headline type condition. Points are probabilities estimated from mixed effects models; error bars are 95% confidence intervals. Point sizes indicate for each cell the proportion of observations.

Perceived accuracy for correctly recognized real news (left panel) was significantly higher when fake news was also recognized as such, $\chi^2(1) = 138.36$, $p < .001$. Additionally, a significant Age × Correction Classification interaction, $\chi^2(1) = 4.32$, $p =.04$, showed that the difference in perceived accuracy based on fake news recognition was greater for younger adults, z ratio = 9.94, $p < .001$, than older adults, z ratio = 7.07, $p < .001$. No other effects were significant, largest $\chi^2(1) = 3.26$, $p = .07$. Furthermore, the model for perceived accuracy of falsely recognized fake news (right panel) indicated a significant effect of Correction Classification, $\chi^2(1) = 6.77$, $p < .001$, that was qualified by a significant Headline Type × Correction Classification interaction, $\chi^2(1) = 5.42$, $p = .02$. When fake news appeared once, perceived accuracy ratings did not differ based on whether corrections were remembered, $t(1250) = 0.32$, $p = .75$. In contrast, when fake news appeared thrice, perceived accuracy ratings were significantly lower when corrections were remembered than when they were not, $t(1258) = 3.50$, $p < .001$. No other effects were significant, largest $\chi^2(1) = 2.58$, $p = .11$.

Discussion

Experiment 1 showed that a lack of fake news exposure and age effects on overall correct and false recognition reflected offsetting effects that depended on detection of and memory for corrections. Three fake news exposures improved detection of corrections in Phase 2 over one fake new exposure for both age groups, but these additional exposures only improved classification of corrections in Phase 3 for younger adults. Moreover, both age groups showed comparable overall detection of corrections in Phase 2, but younger adults classified corrections as such in Phase 3 more accurately than older adults. Despite these age-related memory differences, both groups showed the same conditional recognition patterns. Real news recognition was better when corrections were detected and fake news was recognized as having been corrected, and worse when corrections were detected but fake news was not recognized as such, consistent with previous findings (Kemp et al., 2022b). The benefit associated with detection of and memory for corrections was greater for younger than older adults. A similar benefit was also observed in a reduction in false recognition of fake news that was comparable for both age groups. These benefits join earlier studies suggesting that increasing misinformation accessibility can enhance conflict saliency and integrative encoding when viewing corrections (Ecker et al., 2017; Wahlheim et al., 2020). Here, older adults' smaller real news recognition benefit may reflect poorer integrative encoding, consistent with earlier findings (e.g., Wahlheim, 2014; Wahlheim & Zacks, 2019).

Fake news exposure also led to greater proactive interference shown as diminished real news recognition and increased fake news false recognition when fake news appeared thrice compared to once. This emerged for detected corrections that were not remembered as such. Furthermore, increased exposure to fake news led to higher false recognition when corrections were neither detected nor remembered for older adults, which is consistent with idea that older adults are sometimes more susceptible to interference (Swire et al., 2017a). The negative effects associated with failing to recognize fake news as corrected is compatible with the idea that repeating misinformation can increase memory misattributions based on familiarity (Skurnik et al., 2007). However, the present results are also compatible with the more nuanced view that increasing the accessibility of fake news via repetitions creates more opportunities for co-activation and integration, but when recollection subsequently fails, such accessibility leads to more familiarity-driven memory errors. Although recollection estimates were comparable between age groups, the results of the MPT model showed that repetition of fake news increased parameter estimates of familiarity for older but not younger adults, suggesting that fake news familiarity influenced memory decisions more for older adults.

Finally, overall belief accuracy in Phase 3 reflected in the difference in perceived accuracy ratings for correct real news recognition and false fake news recognition in the correction headline type conditions was comparable across fake news exposures and age groups. Remembering that fake news had been corrected and recognizing it as such was associated with better correct recognition, especially for younger adults. Additionally, remembering that fake news had been corrected was associated with lower beliefs in falsely recognized fake news, but only when that fake news had appeared thrice in Phase 1. Both age groups showed comparable repetition effects of this sort. Collectively, these results show that both age groups can use memory for headline veracity as a basis for beliefs, but the extent to which they did depended on interactions among response types and fake news exposure.

Experiment 2

Experiment 1 provided the first characterization of the effects of repeated fake news exposure and real news corrections on recognition memory and belief accuracy in younger and older adults. One surprising outcome from Experiment 1 was the absence of age-related differences in overall memory accuracy for headlines. Although age-related memory differences are less pronounced when there is strong environmental support, such as in recognition memory tasks (Fraundorf et al., 2019), there were also strong interference resolution and source memory requirements that would typically lead to lower memory performance for older than younger adults (Danckert & Craik, 2013; but see Rhodes et al., 2019). One possible explanation for this lack of age differences is that the online data collection method attracted higher performing and more motivated older adults within their cohort (Ryan & Campbell, 2021). Another possibility is that more prominent age-related differences would have emerged if the final memory task placed greater demands on self-generated retrieval. We designed Experiment 2 to examine whether age-related memory differences would emerge in cued recall instead of recognition and to determine if fake news exposure effects on memory and beliefs replicate using this type of memory test.

Experiment 2 used the same task and data collection methods as Experiment 1 except for the final test procedure. In that phase, headlines appeared with missing details that could be real or fake. Participants were told to 1) recall real news details from Phase 2 and rate their accuracy, 2) indicate if real news details corrected fake new from Phase 1, and 3) if so, recall fake news details from Phase 1. Even if older adults recruited from an online platform are higher performing and more motivated, this final cued recall task may still be more sensitive to age-related susceptibility to interference effects and source confusion. We also expected to observe similar qualitative patterns in conditional recall depending on detection of and memory for corrections as well as interactions of memory and beliefs.

Participants

The stopping rule was to collect usable data from 102 younger and 102 older adults to match Experiment 1. We recruited participants as in Experiment 1 and excluded participants who had completed that experiment. The final sample included 102 younger adults (41 women, 56 men, 5 gender diverse) ages 18-34 years (*M* = 23.60, *SD* = 2.96) and 102 older adults (56 women, 46 men) ages 62-75 years (*M* = 68.17, *SD* = 3.01). We excluded data from 29 younger adults and 48 older adults. Of those participants, 15 younger and 12 older did not return to the second session, nine younger and 22 older did not complete the first session, five younger and 13 older did not complete the second session, and one older adult was exposed to the procedure in the first session multiple times.

Design, Materials, and Procedure

The design, materials, and procedure shared many features with Experiment 1, but there were two key differences. First, participants were told to start the second session 24 hours after completing the first session and were allowed to start up to 49 hours later. The average number of hours between sessions was not significantly different for younger adults (*M* = 26.89, *SD* = 4.73, *Range* = [23.53 – 47.25]) and older adults (*M* = 27.29, *SD* = 5.98, *Range* = [22.95 – 48.80]), $t(202) = 0.34$, $p = .73$. Second, Phase 3 in the second session was a cued-recall test. We shortened the delay between sessions, relative to Experiment 1, to avoid floor effects in cued recall. On the cued recall test, each test item included the picture from the earlier-studied headline above a prompt to recall the missing real news detail that appeared in Phase 2. That detail corrected fake news and affirmed real news that appeared in Phase 1. After attempting to recall a Phase 2 detail by typing a response, participants rated the truthfulness of the detail they recalled using the Phase 1 scale: 1 (Definitely False) to 6 (Definitely True) by clicking a response box on the screen. Following this, they indicated via key press whether real news in Phase 2 corrected fake news from Phase 1 by responding “Yes” (1) or “No” (0). After responding “yes,” they attempted to recall the fake news detail from Phase 1 by typing a response. After responding “no,” they advanced to the next trial.

Results and Discussion

Perceived Accuracy: Phase 1

We compared younger and older adults’ initial perceived accuracy ratings for real and fake news headlines before corrections (Table 7). A model with Age and Headline Type and as fixed effects indicated significant effects of Age, $\chi^2(1) = 9.72$, $p < .01$, and Headline Type, $\chi^2(2) = 67.60$, $p < .001$, and no interaction $\chi^2(2) = 0.36$, $p = .84$. Replicating Experiment 1, younger adults made higher overall ratings than older adults. Both groups made higher ratings for real than fake news, smallest z ratio = 5.39, $p < .001$, and for fake news appearing thrice compared to once, z ratio = 2.68, $p < .001$. These results replicate Experiment 1 and again suggest that younger adults were less skeptical, both groups discerned true from false details, and repeating fake news led to illusory truth.

Correction Classifications: Phases 2 and 3

Phase 2 (Detecting Corrections)

As in Experiment 1, we assessed correction detection in Phase 2 and recognition of fake news corrections in Phase 3 to contextualize downstream effects on memory and beliefs. We again computed probabilities of "yes" responses separately in each phase and made comparisons using separate model with Age and Headline Type as fixed effects (Table 8). The Phase 2 model indicated no significant effect of Age, $\chi^2(1) = 3.32$, $p = .07$, a significant effect of Headline Type, $\chi^2(2) = 2373.07$, $p < .001$, and no significant interaction, $\chi^2(2) = 2.85$, $p = .24$. Both age groups again discriminated fake news corrections from real news affirmations as shown by significantly higher probabilities for both fake news correction types, smallest z ratio = 42.63, $p < .001$. Classification rates did not differ depending on whether fake news appeared thrice or once, z ratio = 1.61, $p = .24$.

We again conducted complementary signal detection analyses of discriminability (d') and bias (c) using separate models with Age and Headline Type as fixed effects for each parameter (Figure 11A, bottom left panel). The model for d' indicated no significant effect of Age, $\chi^2(1) = 0.17$, $p = .68$, a significant effect of Headline Type, $\chi^2(1) = 5.60$, $p = .02$, and no significant interaction, $\chi^2(1) = 0.35$, $p = .56$, showing that discriminability was higher for corrections of fake news than appeared thrice compared to once. In addition, a comparable model for c indicated no significant effect of Age, $\chi^2(1) = 1.39$, $p = .24$, a significant effect of Headline Type, $\chi^2(1) = 5.67$, $p = .02$, and no significant interaction, $\chi^2(1) = 0.35$, $p = .55$.

Phase 3 (Remembering Corrections)

Turning to correction classifications in Phase 3 (Table 8), the model indicated significant effects of Age, $\chi^2(1) = 17.42$, $p < .001$, and Headline Type, $\chi^2(2) = 1756.92$, $p < .001$, that were qualified by a significant interaction, $\chi^2(2) = 22.38$, $p < .001$. Response probabilities were higher for older than younger adults, and for both fake news corrections than real news, smallest *z* ratio = 36.62, $p < .001$, and for fake news that had appeared thrice compared to once in phase 1, *z* ratio = 3.81, $p < .001$. The interaction reflected that of the higher response rates for older than younger adults the false alarm rate to real news topics showed the largest difference. Overall, these results suggest that classification discrimination was better for younger than older adults and following more fake news exposures.

Signal detection parameter estimates compared using the same modeling approach as in Phase 2 verified these discrimination differences. The model for d’ (Figure 11A, bottom right panel) indicated significant effects of Age, $\chi^2(1) = 3.98$, $p = .046$, and Headline Type, $\chi^2(1) = 16.53$, $p < .001$, and no significant interaction, $\chi^2(1) = 0.48$, $p = .49$, showing that younger adults identified topics associated with fake news corrections better than older adults, and such identification by both groups was better when fake news appeared thrice compared to once. In addition, a comparable model for c indicated significant effects of Age, $\chi^2(1) = 25.24$, $p < .001$, and Headline Type, $\chi^2(1) = 16.58$, $p < .001$, and no significant interaction, $\chi^2(1) = 0.51$, $p = .47$, showing that older adults adopted a more liberal response bias than younger adults and that both groups adopted more liberal response biases for corrections of fake news that appeared thrice compared to once. Taken with the same patterns of results in Experiment 1, these results show that when accounting for false alarms to repeated real news headline topics, younger adults showed uniformly better discrimination and more conservative responding than older adults when identifying topics for which fake news was corrected.

Overall Cued Recall: Phase 3

We examined the proactive effects of fake news exposure prior to corrections on subsequent memory accuracy by examining cued recall of real and fake news details in Phase 3. We assessed memory accuracy by comparing correct recall of real news headlines as well as intrusions and correct recall of headlines using separate models for each memory measure with Age and Headline Type as fixed effects.

Correct Recall of Real News

Figure 12D displays correct recall of real news, which refers to when participants recalled the Phase 2 detail. The model indicated no significant effect of Age, $\chi^2(1) = 0.38$, $p = .54$, a significant effect of Headline Type, $\chi^2(2) = 258.78$, $p < .001$, and no significant interaction, $\chi^2(2) = 0.62$, $p = .73$. The pattern paralleled Experiment 1 in showing that repeating real news led to higher recall than correcting fake news, smallest z ratio = 13.89, $p < .001$, and that recall did not differ based on fake news exposure, z ratio = 0.35, $p = .94$. Unsurprisingly, these results show that repetition of real news facilitated memory. More interesting, as in Experiment 1, the lack of fake news exposure effects points to offsetting facilitation and interference effects depending on detection of and memory for corrections.

Intrusions of Fake News

Figure 12E displays intrusions of fake news, which refers to when participants recalled Phase 1 fake news details when trying to recall the Phase 2 real news details. Note that intrusions for repeated real news topics were fake news details that never appeared in the experiment. Those details would have appeared in Phase 1 if those topics had been assigned to one of the fake news headline types. Such responses were therefore intrusions from semantic memory that provide baseline measures of prior knowledge and guessing. The model indicated no significant effect of Age, $\chi^2(1) = 3.22$, $p = .07$, a significant effect of Headline Type, $\chi^2(2) = 256.14$, $p < .001$, and no significant interaction, $\chi^2(2) = 1.68$, $p = .43$. Similar to Experiment 1, intrusions of fake news were significantly higher for both types of corrected fake news headlines than repeated real news, smallest z ratio = 12.29, $p < .001$. In contrast to false recognition in Experiment 1, intrusions were significantly higher for fake news that had appeared thrice than once, z ratio = 4.77, $p < .001$. These results suggest that the expression of proactive interference from additional fake news exposures was greater when the retrieval requirement was to self-generate responses. This was likely because participants were less able to compare details between response options than when real and fake news details were provided by the experiment during recognition testing.

Correct Recall of Fake News

Figure 12F displays correct recall of fake news, which refers to when participants recalled fake news details after indicating that real news corrected the fake news in Phase 2. Note that responses for repeated real news headlines were fake news details that did not appear in the experiment. Such responses were therefore details from semantic memory that provide baseline measures of prior knowledge and guessing. The model indicated no significant effect of Age,

$\chi^2(1) = 0.32$, $p = .57$, and a significant effect of Headline Type, $\chi^2(2) = 1042.98$, $p < .001$, and a significant interaction, $\chi^2(2) = 13.73$, $p < .01$. Compared to younger adults, older adults were more likely to correctly recall fake news details for repeated real news topics, *z* ratio = 3.28, $p < .01$, which may reflect older adults having more prior knowledge of news (Brashier et al., 2017). However, there were no age differences for either fake news correction headline type, largest *z* ratio = 0.41, $p = .41$. Finally, fake news recall was significantly higher when fake news appeared thrice compared to once, *z* ratio = 8.04, $p < .001$. These results are consistent with the other measures in showing comparable fake news exposure effects on memory accessibility for both age groups.

Cued Recall in Phase 3 Conditionalized on Correction Classifications from Phases 2 and 3

Correct Recall of Real News

Following the approach from Experiment 1, we next characterized the offsetting effects of fake news repetitions that depended on detection and recollection of corrections by conducting conditional analyses of real and fake news recall (Figure 13). We first examined the extent to which real news recall depended on participants detecting corrections in Phase 2 and recalling Phase 1 fake news headlines in Phase 3 by conditioning real news recall for the two correction headline types on those variables (Figure 13C). The observation frequencies corresponding to point sizes are displayed in the bottom rows of Table S10, and the complete model results are displayed in Table S11. We used a model including Age, Correction Detection (Phase 2), Fake News Recall (Phase 3), and Headline Type as fixed effects. Below, we highlight key main effects and describe their qualifications by other variables by reporting results from two-way models for each combination of correction detection and fake news recall including fixed effects of Age and Headline Types.

Consistent with prior findings, detecting corrections and subsequently recalling fake news headlines were both associated with better recall of real news headlines, smallest $\chi^2(1) = 95.43$, $p < .001$. The differences associated with detecting corrections and recalling fake news can be seen in Figure 13C by looking between columns and rows, respectively. These effects were qualified by two-way interactions, smallest $\chi^2(1) = 6.94$, $p < .01$. Similar to Experiment 1, when corrections were detected and fake headlines were recalled as such (top left panel), the real news recall benefit was substantially larger when fake news was recalled than when corrections were detected, and real news recall was best when both occurred. The benefit associated with fake news recall was also greater for younger than older adults. Finally, a significant effect of Headline Type indicated that correct recall was higher when fake news appeared once compared to thrice, but this pattern was nominally inconsistent across cells, so interpretative caution is warranted. Similar to Experiment 1, these results indicate that overall recall reflected combinations of enhanced and impaired memory depending on detection of and memory for corrections. However, unlike Experiment 1, the memory impairment observed when detected corrections did not lead to recollection of fake news as being corrected was not selectively greater after additional fake news exposures.

Intrusions of Fake News

We next examined the extent to which intrusions of fake news headlines depended on participants detecting corrections (Phase 2) and remembering corrections (Phase 3) using the same modeling approach as in the previous analyses (Figure 13D). The observation frequencies corresponding to point sizes appear in the bottom rows of Table S10, and the complete model results appear in Table S12.

Also consistent with prior findings, detecting and subsequently remembering corrections were both associated with fewer fake news intrusions, smallest $\chi^2(1) = 62.41$, $p < .001$. There were also significant effects of Age, $\chi^2(1) = 8.53$, $p < .01$, and Headline Type, $\chi^2(1) = 7.74$, $p < .001$, as well as two-way interactions including different combinations of these variables except Age, smallest $\chi^2(1) = 7.82$, $p < .01$. The differences associated with detecting and remembering corrections can be seen by looking between columns and rows, respectively (Figure 13D). When corrections were detected and subsequently remembered (top left panel), the benefit to reductions in intrusions of fake news was greater for younger than older adults, $\chi^2(1) = 8.37$, $p < .001$, and this benefit was comparable following one and three exposures to fake news in Phase 1. In contrast, when corrections were detected but not subsequently remembered (bottom left panel), there were more intrusions when fake news had appeared thrice compared to once in phase 1, $\chi^2(1) = 21.82$, $p < .001$, reflecting larger proactive interference effects following more exposures to fake news. Furthermore, when corrections were not detected but were subsequently remembered (top right panel), there were more intrusions when fake news had appeared thrice compared to once in phase 1, $\chi^2(1) = 4.18$, $p = .04$. The same pattern was observed when corrections were not detected and not subsequently remembered (bottom right panel), showing that there were more intrusions when fake news had appeared thrice compared to once in Phase 1, $\chi^2(1) = 21.31$, $p < .001$. Taken with the conditional real news recall results, these results show that the increase in proactive interference following additional fake news exposure was expressed primarily in fake news intrusions, except when corrections were detected, and fake news was recollected as having been corrected.

Recollection and Familiarity Process Estimates for Phase 3 Recall

The cued recall data of the fake news repetition conditions were fit with the same MPT model from Experiment 1. The three response types analogous to those in Experiment 1 were included (i.e., correct recall of real news from Phase 2, intrusion error of fake news from Phase 1, and recall of never-presented fake news), and omission errors were excluded. Table 9 shows that as in Experiment 1, recollection estimates did not differ based on age or fake news exposure. Inconsistent with Experiment 1, additional fake news exposure credibly increased familiarity in younger (0.14 [0.05, 0.22]) but not older (0.03 [-0.07, 0.12]) adults. However, both age groups showed familiarity estimates far lower than those in Experiment 1. This is perhaps not surprising given that cued recall required self-generation, which led to low observation counts for correct real news and intrusions of fake news. This may have reduced the sensitivity to a fake news repetition effect in older adults or an age effect in familiarity. The model fit of the covariances in the data was also not adequate for either age group. These results should thus be interpreted cautiously.

Perceived Accuracy Ratings (Beliefs) for Recalled Headlines in Phase 3

We next examined differences in perceived accuracy for correction headline types that participants recalled as real news in Phase 3. To assess belief accuracy for headline details, we compared accuracy ratings for correct recall of real news from Phase 2 and intrusions of fake news from Phase 1 (Figure 14, right panel). Belief accuracy was indicated by the degree to which participants perceived correctly recalled real news as more accurate than intrusions of fake news. A model with Age, Response Type, and Headline Type as fixed effects indicated a significant effect of Response Type showing that perceived accuracy ratings that led to correct real news recall were significantly higher than those that led to intrusions of fake news, $\chi^2(1) = 252.25$, $p <$

.001. No other effects were significant, largest $\chi^2(1) = 3.55$, $p = .06$. The complete model results are displayed in Table S13. Participants perceived recalled real news as more accurate than intrusions of fake news to the same degree regardless of the amount of initial fake news exposure.

We further assessed the relationship between memory and beliefs by modelling accuracy ratings separately for each response type conditionalized on memory for corrections in Phase 3 (Figure 15, bottom panels). The models included Age, Headline Type, and Correction Classification as fixed effects. Note that as for similar conditional analyses above, Correction Classification refers to whether fake news was recalled after corrections were identified as such for correct recall of real news and to whether corrections were classified as such following intrusions of fake news. The complete results appear in Table S14.

Perceived accuracy for correctly recalled real news (left panel) was significantly higher when fake news was also correctly recalled at test, $\chi^2(1) = 34.56$, $p < .001$, and for older adults than younger adults, $\chi^2(1) = 4.98$, $p = .03$. No other effects were significant, largest $\chi^2(1) = 0.82$, $p = .37$. Furthermore, the model for perceived accuracy of intrusions of fake news (right panel) indicated a significant effect of Correction Classification, $\chi^2(1) = 10.60$, $p = .001$, that was qualified by a significant Age × Correction Classification interaction, $\chi^2(1) = 19.76$, $p < .001$. Younger adults showed no significant difference in perceived accuracy depending on whether corrections were remembered, $t(797) = 1.14$, $p = .25$, whereas older adults indicated that intrusions were significantly less accurate when they remembered corrections compared to when they did not, $t(809) = 5.38$, $p < .001$. No other effects were significant, largest $\chi^2(1) = 1.24$, $p = .27$.

Discussion

Following corrections of fake news, overall correct recall for real and fake news details as well as intrusions of fake news details were comparable between age groups. This lack of age-related memory differences conceptually replicates the findings in Experiment 1. Unlike Experiment 1, additional fake news exposure increased intrusions comparably for both age groups, suggesting that the self-generated retrieval task here was more sensitive than recognition to proactive interference effects. This additional interference can be accounted for by dual-process models holding that such errors reflect familiarity-based misattributions (Jacoby, 1999). Consistent with this, parameter estimates derived from an MPT model showed a greater contribution of familiarity to cued recall following repeated fake news exposure for younger adults. However, these results are somewhat ambiguous due to difficulties with model fitting resulting from low response counts. Moreover, the lack of evidence for greater interference susceptibility in older adults after repeated fake news exposure is inconsistent the dual process view that older adults have impaired recollection.

These age-related similarities in memory performance may also account for older adults' ability to detect fake news correction as well as younger adults. If such detection is based partly on memory for conflicting information, then older adults preserved memory for initially encoded fake new may have supported these judgments. However, as in Experiment 1, older adults were subsequently less able to remember which headline topics were earlier corrected, thus revealing a type of deficit in associative memory for veracity information. Also as in Experiment 1, the primary age-related memory differences were observed when considering how memory for headline details varied based on whether corrections were detected and remembered. Younger adults showed greater benefits to correct recall of real news and reductions in intrusions of fake news associated with memory for corrections. This suggested that younger adults were better able to encode associations between real and fake news details. However, repeated fake news exposures did not increase these age differences.

Finally, conceptually replicating Experiment 1, overall belief accuracy in Phase 3 reflected in the difference in perceived accuracy ratings for correct real news recall and intrusions of fake news in the correction headline type conditions was comparable across fake news exposures and age groups. Although correctly recalling fake news in Phase 3 was associated with more accurate beliefs in correctly recalled real news, this effect was also comparable across fake news exposures and groups. However, an age difference emerged in conditional analyses of belief accuracy for intrusions of fake news. Older adults showed more accurate beliefs when they remembered corrections, while younger adults did not enjoy those same benefits. These results collectively suggest that both age groups can use memory as a basis for beliefs. However, the extent to which they did depended on response type, as in Experiment 1, but not fake news exposure, which contrasts with Experiment 1. These across experiment differences indicate that the distinction between recognition and cued recall also determines how memory representations are accessed when evaluating headline accuracy.

General Discussion

Two experiments characterized the effects of correcting repeated fake news on memory for real news details and perceived accuracy of retrieved details in younger and older adults. Both age groups showed consistent patterns in overall recognition and cued recall accuracy for learned details, but fake news repetition effects varied across measures. Fake news repetition did not affect recognition or recall of real news, nor did it affect false recognition of fake news;

however, fake news repetition increased correct recognition, correct recall, and intrusions of fake news. The lack of overall fake news repetition effects reflected offsetting enhancement and impairment that depended on how often participants remembered earlier-detected corrections as such and the corrected fake news details. Repeating fake news improved remembering that fake news was corrected and retrieving the fake details themselves, which were together associated with enhanced memory for real news. However, when detected corrections led to failed recognition and recall of fake news as well as failed remembering that topics had been corrected, fake news repetition impaired real news recall and increased fake news intrusions. Despite the similar patterns between age groups, closer inspection revealed that the real news recall enhancement associated with remembering fake news corrections was smaller for older than younger adults. This likely reflected older adults' less precise memory for corrections shown by more inaccurate classifications of real news details that did not correct fake news. Note that this age difference was subtle enough to preclude detection of an overall memory difference.

Both groups reported greater perceived accuracy in overall correct recognition and recall of real news compared to false recognition and intrusions of fake news that did not depend on fake news exposure. The higher perceived accuracy for correct recognitions and recalls was even greater when participants accurately recognized and recalled fake news details as such. This enhancement in perceived accuracy was greater for younger than older adults in recognition and comparable between age groups in cued recall. A more complex picture emerged in accuracy perceptions for false recognitions and intrusions. Remembering corrections was associated with lower perceived accuracy in false recognitions, but only when fake news was earlier repeated. Remembering corrections was also associated with lower perceived accuracy in intrusions of fake news—but only for older adults—and this pattern did not differ based on prior fake news exposure. Collectively, these findings underscore the complex interplay of age, memory, and perceived accuracy that emphasizes the need to consider how task-specific retrieval conditions influence the precision of remembering corrections of fake news that contribute to the updating of memory and perceived accuracy.

The present findings have implications for the controversy regarding the effects of repeating misinformation before or during corrections on subsequent memory, beliefs, and reasoning. Two prominent views have been proposed to explain these effects. The familiarity-backfire view proposes that repeating misinformation may reduce correction efficacy by increasing familiarity and processing fluency (Autry & Duarte, 2021; Schwarz & Jalbert, 2020). Conversely, integrative-encoding view proposes that repeating fake news may enhance correction effectiveness by promoting conflict saliency and the co-activation of both the misinformation and its correction (Ecker et al., 2017). Both views are compatible with a dual-process perspective, which emphasizes that memory accuracy for corrected information varies depending on whether retrieval is familiarity-based, relying on a general feeling of memory strength, or recollection-based, involving the retrieval of contextual details, which may include associations between true and false information (e.g., Wahlheim et al., 2020).

Our study revealed evidence compatible with both perspectives. Critically, whether familiarity-based influences of fake news exposure exerted unwanted effects on memory and perceived accuracy depended on age, task features, and memory for corrections. Repeating fake news provided more opportunities for improving memory updating by increasing the extent to which detected corrections and fake news details were later recollected. This finding is compatible with the integrative encoding perspective that recollection-based retrieval can grant access to relationships between true and false information established while corrections are being processed. However, repeated fake news exposure impeded memory updating more when detected corrections and fake news details were not later recollected. This is compatible with the view that familiarity-based errors arise when they are not opposed by recollection of criterial details, such as the information source. This mixture of enhancement and impairment created by fake news repetitions corresponds with related research showing that presenting fake news reminders with corrections dramatically improves subsequent memory for corrections and recollection-based retrieval that fake news had been corrected (Kemp et al., 2022ab; Wahlheim et al., 2020). Moreover, these findings are compatible with the Memory-For-Change framework, which proposes that retrieving outdated information during while encoding stimuli with similar but not identical features can subsequently improve memory when the stimulus changes are recollected but impair memory when such changes are not recollected (Jacoby et al., 2015; Wahlheim & Jacoby, 2013).

The present study also illuminates the effects of repeated fake news exposure on memory and belief accuracy in both younger and older adults. Dual-process theories propose that as people age, their ability to recollect specific details declines, leading them to rely more on familiarity during retrieval (Ayers & Reder, 1998; Jacoby, 1991; Yonelinas, 2002). We therefore assumed that the increase in familiarity from repeating fake news could more detrimentally affect older adults' memory accuracy (cf. Jacoby, 1999). Overall memory accuracy was comparable for both groups, but older adults remembered fewer corrections and enjoyed fewer associated benefits to memory accuracy, especially when the task required self-generation of news details. These findings align with work showing that older adults were more prone to erroneously remembering myths as facts after repeated retractions, especially following a three-day retention period (Skurnik et al., 2005; but see Swire et al., 2017a). However, the present age differences in memory for corrections were too subtle to impact overall memory. This lack of robust age differences may have resulted from the naturalism of materials, as older adults may have leveraged their intact semantic memory (Park et al., 2002) when learning the headline content. Indeed, older adults are more inclined to engage with news content than younger adults (Brashier & Schacter, 2020) and have been shown to use prior knowledge to prevent making inaccurate judgements induced by repetition (Brashier et al., 2017).

One unexpected finding was that fake news repetitions did not lead to differences in overall recognition or recall of real news details. This finding diverges from studies showing that repeating misinformation increased its influence on inferential reasoning compared to a single exposure (Ecker et al., 2011). This discrepancy may have arisen because more repetitions were needed to elicit detectable differences. Future work could examine whether correction efficacy on memory updating varies across a wider range of fake news repetitions (cf. Udry et al., 2022; Experiment 2) and interactions with different retention intervals. The latter task feature is important to consider because recollection decreases over time for both younger and older adults. Thus, any advantage that younger adults may have enjoyed could have been diminished by the two-day retention interval here.

The present study also contributes to the nascent literature on the interplay between memory and beliefs in misinformation corrections. Previous work in this area mainly focused on correction effects on inferential reasoning with less emphasis on the role of memory (for a review, see Lewandowsky et al., 2012). However, it is well established that memory serves as one basis for beliefs (Begg et al., 1992; Berinsky, 2017; Kowalski & Taylor, 2017; Newman et al., 2022; Swire-Thompson et al., 2023; but see, Collier et al., 2023). In our prior work using similar fake news correction paradigms, we showed that memory for corrections was associated with more accurate beliefs—defined as the extent to which perceived accuracy for correct recall of real news details was higher than for intrusions of fake news details—especially when corrections included veracity labels (Kemp et al., 2022a; Wahlheim et al., 2020). The present findings replicate that work in showing that memory for corrections was associated with higher perceived accuracy of recalled real news and extend those findings to recognized real news, regardless of fake news exposures. These benefits were slightly greater for younger than older adults in recognition, but this age difference did not extend to cued recall. Closer inspection of the recognition results suggest that this effect was mostly driven by the condition including one fake news exposure, despite the absence of a statistical interaction. We are therefore uncertain about the reliability of this age difference.

In contrast to the similarities in perceived accuracy for retrieved real news, the patterns of perceived accuracy for false recognition and intrusions of fake news were inconsistent between experiments. In Experiment 1, both age groups showed that remembering that fake news had been corrected was associated with lower beliefs in falsely recognized fake news, but only when that fake news had appeared three times earlier. However, in Experiment 2, the association between remembering that fake news had been corrected and lower beliefs in fake news intrusions was comparable across fake news exposure conditions. These findings show that the moderation of the relationship between memory for headlines and beliefs by fake news repetitions becomes more prominent when retrieval is externally prompted rather than self-generated. Another difference was that in Experiment 2, older adults exhibited more accurate beliefs when they remembered corrections, but younger adults did not enjoy those same benefits. Given that we did not a priori assert any theoretically motivated hypotheses about these differences, we can only speculate here about their potential causes. For false recognition, it is possible that the chosen headlines that were initially repeated more often induced a stronger sense of familiarity that cast doubt on the veracity of the headlines. Moreover, for intrusions of fake news, it is possible that both age groups used different decision criteria for outputting headline details under conditions of uncertainty. Older adults may have been more liberal in generating responses to increase the overall memory quantity scores because they could subsequently express less certainty in response veracity.

The present study provides valuable insights into how repeated fake news exposure affects memory and perceived accuracy for retrieved real news in both younger and older adults, but there are limitations. First, to recruit broadly across the US population, we leveraged an online data collection platform. However, our resources constrained the time available for testing sessions and, consequently, our ability to fully characterize our participants. As noted above, we did not expect such strong similarities in memory accuracy between age groups given what we know about age differences in episodic memory. Although older adults may just be better at remembering news stories than more basic memory task materials such as unrelated word pairs, an alternative explanation is that the online recruitment method attracted older adults who were more motivated (Ryan & Campbell, 2021). Moreover, it is worth noting that our sample of older adults may not be entirely representative of their age cohort. Prolific tends to draw from a pool of generally healthy, technologically proficient, and well-educated individuals (Turner et al., 2020). Additionally, we did not directly assess our participants' cognitive abilities, and thus, it remains a possibility that age-related repetition-induced memory differences, similar to those observed in Jacoby (1999), could have emerged had our older adults been more in line with the average characteristics of their age group. Despite these potential limitations, it is worth emphasizing the merit of employing online data collection methods. This approach better mirrors the way people engage with news in their daily lives, enhancing the real-world applicability of our findings. Our findings are also encouraging in showing that some groups of older adults can perform almost as well as younger adults in an everyday memory task.

Furthermore, in Experiment 2, the MPT model that we fit to the cued recall responses did not converge as expected. Consequently, we were unable to confidently interpret the parameter estimates that were meant to help us more precisely evaluate the dual-process interpretation of our findings. Additionally, our choice of the MPT model was specific to testing the assumptions of one theoretical framework, and there exist several other theories that could have been chosen and subjected to testing. We purposely selected a dual-process theory (Jacoby, 1991, 1999) to motivate our MPT model due to its suitability for assessing the influence of both the recollection of headline veracity and the acontextual familiarity of headline topics on the final recall of real news and based on its suitability in our prior work (Kemp et al., 2022a). However, a future study could be dedicated to comparing the fits of all available models to these data. Given the already substantial contributions offered by the plethora of analyses here, we defer such comparisons to future work.

Concluding Remarks

In conclusion, our study contributes to the growing body of literature on how repeated exposure to fake news influences memory and belief accuracy for fake news corrections in younger and older adults. Neither age nor fake news repetitions affected overall recognition or cued recall of correct news or false alarms to fake news, but repeated fake news was both better recalled and intruded more in cued recall for both age groups. Conditional analyses showed that repeating fake news was especially harmful to memory and belief accuracy when corrections were detected as such, and fake news was not later remembered. Importantly, fake news repetitions provided more opportunities to observe the benefits of remembering corrections but also led to more impairment when corrections were not remembered. Older adults were less able to enjoy the benefits of remembering corrections, but this deficit was too subtle to produce detectable difference in overall performance. Together, these results highlight the key role of recollection-based retrieval of real news corrections following their initial detection on subsequent memory and belief accuracy for real news details. These findings provide a start towards establishing a foundation for research aimed at identifying effective methods for mitigating the detrimental impact of repeated misinformation exposure in an age-diverse society.

CHAPTER V: INTEGRATIVE DISCUSSION

The goal of this dissertation was to integrate work from the continued influence effect, illusory truth effect, and episodic memory literature to understand how fake news influences the efficacy of corrections on memory and beliefs. Given the prevalence of fake news in the real world and its impact on both individual and societal well-being, it is critical to seek effective correction methods, as well as to illuminate the underlying mechanisms for such effects. To this end, the current research program contributes to the literature by revealing that increasing accessibility to fake news through reminders and retrieval bolstered memory and belief accuracy, while repeated exposure led to neither improvements nor impairments. Conditional analyses across the three papers showed that increasing accessibility to fake news provided more opportunities for both the fake news and correction details to be integrated together, and memory and belief accuracy was greater when both details were later recalled. However, when the correction details were not recalled, increasing accessibility to fake news resulted in impairments to memory and belief accuracy, thus emphasizing the importance of recollection-based retrieval.

In the first empirical paper (Kemp et al., 2022a), the efficacy of reminder-based corrections was compared against veracity-labeled real news corrections without reminders (Experiment 1) and veracity-labeled fake news on its debut (Experiment 2). The results of Experiment 1 revealed how conflict saliency and integrative encoding separately contribute to memory and belief accuracy, and that the benefits conferred by reminder-based corrections were more attributable to the enhanced integrative encoding that supported recollection-based retrieval. The results of Experiment 2 further highlighted the critical role of integrative encoding for reminder-based corrections, as well as showing how labeling fake news on its debut benefits belief accuracy but not memory accuracy. The goal of the second empirical paper (Kemp et al., 2022b) was to investigate the impact of self-initiated detection of corrections and the retrieval of fake news details during the presentation of corrections on downstream memory accuracy for headlines. The conditional analyses presented evidence that could be partially explained by the familiarity backfire view and the integrative encoding account. However, a more coherent understanding of the results was provided by the MFC framework which not only incorporates elements from both perspectives but also emphasizes the moderating role of recollection-based retrieval. Specifically, it was revealed that retrieving fake news details during corrections improved memory accuracy when fake news was later recollected but impaired memory accuracy when fake news was not later recollected. In the third empirical paper (Kemp et al., 2023), the effects of repeated fake news exposure before corrections influence memory and belief accuracy for younger and older adults were examined. Findings showed that neither age nor repeating fake news impacted overall memory accuracy for real news headline details, but repeating fake news did increase false alarm and intrusion errors. Again, conditional analyses revealed that repeating fake news provided more opportunities for integrative encoding but also more impairment when corrections were not remembered. Further, older adults were less able to enjoy the benefits of remembering corrections, but this did not lead to overall differences in memory performance compared to younger adults.

In the following sections, I discuss the theoretical and practical implications that this work has for our comprehension of the impact of fake news exposure on correction efficacy and age-related differences in this context. Then, I will discuss the importance of examining the association between memory and beliefs in the context of fake news exposure effects. Finally, I will conclude with a section dedicated to exploring the strengths and limitations of this research program. Throughout each section of this integrative discussion, I offer insights into potential directions for future research.

Theoretical Implications

Familiarity Backfire Versus Integration

This research program has significant theoretical and practical implications. Firstly, the present results are relevant for the theoretical debate about whether corrections should repeat misinformation details. According to the familiarity backfire account, the act of repeating misinformation details while correcting can increase the perceived accuracy of the information (Skurnik et al., 2007; as cited in Schwarz et al., 2007). The argument posits that repetition increases familiarity (Schwarz et al., 2007) and the ease of processing (Reber & Schwarz, 1999; Unkelbach, 2007), potentially leading to misattributions of accuracy. In contrast, integration accounts argue that repeating misinformation may enhance correction efficacy by increasing conflict saliency and promoting integrative encoding (Wahlheim et al., 2020; for a review, see Ecker et al., 2022). The convergence of results observed here more closely aligns with integration accounts than the familiarity backfire account, in showing that reiterating fake news details via reminders and retrieval improved overall memory and belief accuracy.

More evidence in support of integration accounts came from the conditional analyses, showing that improvements in memory and belief accuracy were primarily driven by the repetition of fake news promoting the integration of memories to support the recollection of corrections. However, the conditional analyses also showed that memory and belief accuracy were impaired when people could not remember that corrections occurred. These findings align with the dual-process perspective, suggesting that when unopposed by controlled recollection, the familiarity induced by the repetition of misinformation can lead to memory misattribution errors (Ecker et al., 2011; Swire et al., 2017a). This nuanced understanding sheds light on the conditions under which familiarity errors may manifest. More broadly, the mixture of effects that depends on recollection accords with the MFC framework (Wahlheim & Jacoby, 2013; Wahlheim et al., 2020) emphasizing the necessity of both successfully integrating and recollecting corrections.

Fake News Exposure on Older Adults

The present study also contributes to the nascent research on aging and the effects of fake news exposure. Few studies in the continued influence effect literature (Guillory & Geraci, 2010; Swire et al., 2017a) have included older adults, leaving age-related differences relatively unknown. Here, I found that there were no age-related differences in overall memory performance for real news headlines following repeated exposure to fake news. This finding challenges the prevailing notion that older adults are more susceptible to memory interference from previously learned information due to age-related deficits with recollection (Jacoby, 1999), inhibition (Hasher & Zacks, 1988), and encoding (Spencer & Raz, 1995). That said though, older adults remembered fewer corrections and derived fewer associated benefits to memory accuracy. This may reflect older adults' binding deficits disrupting integrative encoding when both the fake and real news were coactivated (for a meta-analysis, see Old & Naveh-Benjamin, 2008); and/or recollection-deficits impairing their ability to recall integrated memory representations (Wahlheim, 2014). These findings emphasize the need for future research to concentrate on developing correction strategies that specifically target age-related deficits. Nonetheless, the present work is encouraging because it shows that older adults are not more susceptible to fake news, even under circumstances that should facilitate age differences (cf. Swire et al., 2017a).

Improving Fake News Interventions in Real-World Contexts by Enhancing Conflict Saliency and Integrative Encoding

This research program also provides practical insights for addressing fake news in real-world contexts, emphasizing the need for interventions to enhance conflict saliency, integrative encoding, and recollection-based retrieval. Drawing from the data, social media platforms and policymakers could implement a system that includes the provision of fake news reminders, veracity labeling of information, and prompting users to recall specific details of encountered false information. However, while the benefits of such interventions have been observed in an experimental setting, these benefits may not transcend to the real world due to factors like divided attention (Sanderson et al., 2022), engagement metrics influencing the perceived accuracy of the information (Butler et al., 2022), the sheer amount of accurate information, potentially impeding the identification of fake news. Recent work in the continued influence effect literature has started to use social media simulations (Butler et al., 2023) designed to mimic real-world experiences, thereby enhancing ecological validity. Future research should capitalize on this simulation to explore how decreasing attention by incorporating another task, manipulating the valence (i.e., likes, dislikes) and/or quantity of engagement metrics, and manipulating the ratio of true and false information affects the efficacy of fake news reminders, veracity labels, and retrieval.

Beyond the interventions directly explored here, it would be useful for future work to consider alternative interventions that may be used in conjunction with fake news reminders to enhance conflict detection and integrative encoding. One avenue worth exploring involves manipulating perceptual characteristics of correction details that conflict with fake news information to explore their impact on subsequent memory accuracy. Past studies suggest that manipulating perceptual features, such as displaying information in larger fonts (Luna et al., 2019) or distinct colors (Santos et al., 2019), can improve recall as they are perceived as more important and attention-grabbing. Therefore, if conflicting details in both the fake news reminder and correction are presented in larger fonts and/or different colors, this may potentially facilitate conflict detection and integrative encoding as the information is encoded more efficiently and becomes more salient in memory.

Along the same vein, it would be interesting to explore how manipulating the degree of perceptual similarity between fake news and its correction influences subsequent memory accuracy. The stimuli used in this study were inspired by work from the episodic memory updating literature such that there were minor differences between the details in the fake news and correction headlines. This was intentionally done so that during the final test phase a prompt could be answered with details from either headline (i.e., false or real version). However, in everyday life, fake news and real news are rarely worded in the same perceptual manner. This prompts the question of whether higher or lower perceptual similarity is more conducive for integration and subsequent recollection. On one hand, higher perceptual similarity may boost later memory accuracy by triggering more reminders of the fake news details than lower perceptual similarity, in turn facilitating conflict detection and integration. On the other hand, high perceptual similarity may also compromise memory accuracy more than lower perceptual similarity if individuals fail to recollect that a correction occurred, as the representations lack distinctiveness. These ideas remain speculative, underscoring the necessity for future empirical testing, as the outcomes hold practical implications for social media platforms aiming to correct fake news.

Challenges and Considerations of Enhancing Conflict Saliency and Integrative Encoding

As mentioned briefly above, interventions aimed at boosting conflict saliency and integrative encoding are not considered "safe" strategies—they have their potential drawbacks. With respect to veracity labels, research has shown that fact-checkers flagging fake headlines as false can unintentionally lead individuals to perceive unflagged fake headlines as more accurate (Pennycook et al., 2020). Another study has found that warning people about potential misinformation may induce skepticism to the extent that even true information is seen as less accurate (Clayton et al., 2020). While fact-checking flags and general warnings are not inherently the same as veracity labels, the same consequences may occur. Lastly, adding veracity labels to every piece of information is not only laborious for social media companies but also potentially cognitively taxing for individuals. The increased cognitive load may have adverse effects on one's ability to remember both the veracity tag and content, as well as hinder analytical thinking to discern truth (cf. Pennycook & Rand, 2019).

With respect to the downsides of reminders and retrieving fake news, we know from this research program that in the absence of recollection, memory and belief accuracy were impaired, presumably due to increased accessibility to fake news without critical veracity information. This implies that under conditions that impede controlled retrieval processes and promote reliance on automatic retrieval processes, such correction strategies may do more harm than good. In the real world, these correction strategies may prove counterproductive in instances where people encode the fake news details but not the correction details due to divided attention, fatigue, or the inability to identify that a correction has appeared. Alternatively, they may serve to be counterproductive when people encode both the fake news and correction details but fail to remember the source, possibly due to the amount of time that has elapsed and the interference from intervening events. Recognizing these challenges is crucial for refining correction strategies to effectively address misinformation in various everyday contexts at scale.

Association Between Memory and Beliefs

Prior to this research program, most studies in the continued influence literature examining the effects of misinformation exposure focused on inferential reasoning or belief accuracy (for a review, see Pennycook & Rand, 2021), with only a few recent studies exploring its impact on memory accuracy (Ecker et al., 2011; Wahlheim et al., 2020). I believe that examining how misinformation affects memory and belief accuracy is important given that memory is considered to be a basis for beliefs (Berinsky et al., 2017; Kowalski & Taylor, 2017; Newman et al., 2022). Hence, two empirical papers within this research program investigated the relationship between memory and belief accuracy. The results revealed that accurate recall correlated with higher perceptions of accuracy, while incorrect recall was associated with elevated perceptions of truth. Similar findings have been reported by studies that directly explore the role of memory in belief regression (Swire-Thompson et al., 2023; Wahlheim et al., 2023). For example, Swire-Thompson et al. (2023) discovered that beliefs corrected after exposure to misinformation regressed more when participants reported not remembering the statements being corrected. Likewise, an unpublished study by Wahlheim et al. (2023) showed that belief regression over a week and month was reduced most when reminder-based corrections were provided, and such effects were attributed to those corrections enhancing recollection.

While our findings provide an initial exploration of the association between memory and beliefs, future studies should incorporate alternative measures and use different methodologies to measure beliefs which could deepen our understanding of its association with memory. In the first and third papers of this research program, belief accuracy was assessed during the final test phase using a single-item measure (i.e., Rate your belief that the response you recalled was true). However, a review article (Swire-Thompson et al., 2020), underscored the limitations of single-item measures in measuring beliefs due to their susceptibility to random measurement error, inability to assess reliability, and insufficient capacity to capture the breadth of beliefs. To address these concerns, future studies should follow the advice from this dissertation and consider adopting multi-item measures. These measures could include several statements or questions related to belief in a given news story topic, with participants rating their agreement or disagreement with each statement. The more we can accurately capture beliefs, the more adeptly we can assess how beliefs are associated with memory.

Subsequent studies might consider incorporating additional measures, such as assessing people's level of confidence in the details that they remembered (Guillory & Geraci, 2010; Dobbs et al., 2023). Confidence judgments are valuable as they serve as an index of metacognitive awareness. In the first and third empirical papers included here, it is theoretically possible that participants erroneous belief judgments stemmed from overconfidence or underconfidence in the accuracy of the details that they recalled. Moreover, future studies could introduce an open-ended question (e.g., What are the reasons for this belief rating?), prompting participants to articulate the rationale behind their responses. For example, participants could indicate that they "remembered" seeing the information previously, implying that memory served as a basis for their belief response. As will be discussed later, this self-report measure may reveal that other factors beyond memory exert influence on participants' responses (e.g., experiment instructions, worldview, stimuli images).

In addition to diversifying outcome measures, future work could adopt different experimental paradigms. When belief accuracy was assessed in the first and third empirical papers, it was measured through participants' perceptions of the real news headlines that they recalled, which may or may not have corrected earlier fake news. As a result, we cannot address whether corrections also improved belief accuracy for the fake news headlines. It is theoretically plausible that participants might perceive the real news correction headline as accurate but still maintain their belief in the fake news headline. To empirically test this, participants could rate their belief in the fake news headline both before and after corrections, with the fake news headlines appearing identically in both of those phases. Corrections would be deemed ineffective if participants reported the same or higher belief accuracy post-correction compared to pre-correction. It's essential to note that the effectiveness of corrections also hinges on participants' initial beliefs; it could be the case that participants perceived the fake news as false even before corrections.

Recent studies have started to use this test-retest paradigm for correction strategies, including fake news reminders (Wahlheim et al., 2023; Swire-Thompson et al., 2023), but none have explored the impact of fake news retrieval and repeated exposure on belief accuracy. I would speculate that the process of retrieving fake news would lead to similar outcomes on memory and belief accuracy as fake news reminders observed in Wahlheim et al. (2023). In that paper, promoting integrative encoding led to a reduction in belief in misinformation on an immediate test, especially when individuals could remember that a correction had taken place. However, as with fake news reminders, when the correction is not later remembered, fake news retrieval may increase fake news believability. With respect to repeated exposure to fake news, several potential outcomes may emerge. First, the correction could effectively diminish misinformation beliefs, but the effect's magnitude might be diminished due to the repetition of the fake news increasing its believability. Unlike fake news reminders and retrieval, the temporal space between repeated exposure and the correction might solidify belief in the fake news, rendering it harder to correct regardless of whether they remember the correction (Lewandowsky et al., 2012). Alternatively, the fake news belief might become so deeply ingrained that the correction becomes entirely ineffective, irrespective of whether the correction is remembered. Lastly, the repetition of fake news increases its beliefs and memorability, to the extent that participants can detect conflict, integrate the information, and adjust their beliefs accordingly. However, as mentioned earlier, the benefits of integrative encoding are contingent on the subsequent use of recollection-based retrieval.

Strengths and Limitations

The present body of work has several strengths and limitations that should be discussed. Firstly, one strength is that a cognitive lens was used in this research paradigm to advance our understanding of how different manipulations, specifically designed to impact aspects of episodic memory, contribute to the overall efficacy of corrections. Focusing on the role of memory in the context of misinformation and correction strategies is important for several reasons. Memory plays a central role in shaping our beliefs and decision-making processes. When individuals encounter information, their cognitive processes determine how that information is encoded, stored, and later retrieved. In the case of misinformation, people may inadvertently integrate false details into their memory, contributing to the persistence of inaccurate beliefs. By isolating memory, researchers can untangle the specific cognitive processes at play, identifying key mechanisms that influence the success of correction interventions. This targeted focus allows for a more precise understanding of how corrections interact with memory processes, enabling the development of strategies that align with the inherent workings of memory.

Our findings demonstrate that corrections are effective at improving memory and belief accuracy when they promote both integration and recollection-based retrieval processes. This aligns with previous research theorizing that corrections are relatively ineffective due to integration, selective retrieval, and/or working memory failures (for a review see Ecker et al., 2022). Additionally, the present work contributes to the literature by showing that corrections can be successful under conditions that promote both integration and recollection-based retrieval processes. The present work also compared the efficacy of different correction strategies and provided theoretical rationale to elucidate why one strategy may outperform another. While these findings mark a substantial stride forward, they also underscore the importance of further exploration into the memory mechanisms shaping correction efficacy. To practically recommend memory-based strategies (i.e., fake news reminders, fake news retrieval), it would be important to explore the various ways recollection-based retrieval can further be supported to improve their efficacy in both the short-term and long-term.

Moreover, the cognitive lens not only underscores the importance of memory but also advances our comprehension of the intricate association between memory and beliefs. While prior literature has hinted at this association (Newman et al., 2022; Schacter, 2022), few studies have explored this in the context of correcting fake news. Our findings align with this limited previous work, showing that memory indeed plays a role in perceptions of accuracy for headlines (Wahlheim et al., 2020; Swire-Thompson et al., 2023). This is encouraging news as it suggests that focusing on correction strategies that enhance memorability may also achieve a dual benefit by positively influencing beliefs. These insights lay a foundational understanding of the association between memory and beliefs, providing a basis for scholars to further explore the specific conditions under which this association holds.

While it is important to deepen our understanding of the cognitive mechanisms influencing correction strategies, the current research program does not explicitly consider the impact of social mechanisms on observed findings, except for the second empirical paper. In that paper, I explored if varying peer agreement about the veracity of fake news headlines would affect later encoding and retrieval memory processes. While this social norm manipulation did not yield the expected outcomes, the inclusion of both social and cognitive factors in the same study is important given that information is rarely absorbed in a vacuum. Information consumption in the real world is inherently embedded in social contexts, where individuals encounter information of varying accuracy through interactions with friends on social media and in daily conversations.

Another social factor that could be considered in future studies alongside the role of memory is how the misinformation and correction align with an individual's belief system, commonly referred to as their worldview (for a review, see Swire-Thompson et al., 2020). Existing research suggests that correcting misinformation in line with an individual's worldview might render the correction less effective or, in some cases, backfire by intensifying belief in the misconception (Ecker & Ang, 2019; Nyhan & Reifler, 2010; though see Wood & Porter, 2019; Guess & Coppock, 2020). Notably, such effects have been observed even when participants remember that a correction was issued (Lewandowsky et al., 2005). Throughout this research program, explicit instructions were provided during the correction phase, emphasizing that the to-be-presented headlines were verified by fact-checking websites, offering participants ample reason to believe the information. However, it is possible that if the corrective information did not align with their worldview, or if the fake news did align with their worldview, they would be unlikely to accept the correct information as true regardless of whether they could accurately remember it. I expect that the potential effects of worldview may have been limited in the current studies given that the headlines depicted fairly partisan-neutral content. Future studies could delve deeper into this by conducting follow-up studies with the same experimental design but with a more polarized stimulus set that is more conducive to eliciting backfire effects.

Another social factor that warrants attention when considering the efficacy of memory-based corrections is the credibility of the misinformation and correction sources. In this research program, news headline stimuli were presented without a source reference in an attempt to isolate the effects of various memory-based correction methods. However, in everyday life, news headlines are presented by a source (i.e., publisher, media outlet). Previous research shows that corrections tend to be ineffective when a misinformation source is deemed credible (for a meta-analysis, see Walter & Tukachinsky, 2020), and the correction source lacks credibility (Connor Desai et al., 2020; O'Rear & Radvansky, 2020). Moreover, people are inclined to perceive sources as more credible when the sources align with their worldviews (Chaiken & Maheswaran, 1994). For instance, political partisans are more likely to believe misinformation when it originates from politically congruent rather than incongruent or neutral sources (Swire et al., 2017b; Reinero et al., 2023). With this in mind, it would be intriguing to investigate how political congruency with the information source impacts the efficacy of corrections featuring fake news reminders. This correction strategy may be ineffective when fake news is presented from a politically congruent source, but the correction is presented from a politically incongruent source. This potential ineffectiveness could be attributed to individuals paying more attention and encoding information better when it comes from sources congruent with their political beliefs while disregarding information from non-congruent sources.

In summary, while a cognitive approach provides a solid foundation, future work should adopt an interdisciplinary perspective that integrates cognitive and social factors. Specifically, exploring how the correction strategies emphasized in this research program operate when considering agreement with peers using engagement metrics, as well as the congruency of the content and source with one's worldview. An interdisciplinary approach is important for identifying the most effective strategies for addressing the intricate and multifaceted nature of fake news at scale.

Secondly, this research program primarily focused on assessing post-correction memory and belief accuracy. While this is foundational, it is essential to recognize that other outcomes are also necessary to examine. Future work could explore how these memory-based correction strategies influence the act of sharing on social media websites. Interest in fake news sharing intentions has become a burgeoning area of research (for a review, see Pennycook & Rand, 2021), with work showing that many people are willing to share false information despite perceiving it to be inaccurate (Pennycook et al., 2020; 2021). Consequently, someone might accurately remember both the fake and real news headlines but still choose to share the fake information, suggesting ineffectiveness in the memory-based correction strategy within our data. Additionally, it could also be possible that in the absence of memory for the correction, participants may be more willing to share the fake news later on because they are not paying attention to its veracity and do not have a memory basis to discern its veracity.

Expanding our focus beyond sharing intentions, it would also be valuable to examine how memory-based correction strategies influence attitudes and behavioral intentions related to critical real-life issues, such as adherence to medical advice (e.g., COVID-19), political voting preferences, and territorial conflicts (e.g., Israel-Palestinian conflict, Ukraine war). Although some work has touched upon these topic areas (Cantarella et al., 2023; de Saint Laurent et al., 2022; Greene & Murphy, 2021; Swire et al., 2017b; Reza & Sunvy, 2023), there remains a gap in exploring them in combination with memory-based corrections. As mentioned earlier, these memory-based correction strategies may be less effective when using more politicized and polarized content because belief accuracy can be driven by people's social identity and worldview.

Related to this point, the reminder-based corrections used in Kemp et al. (2022a) and other studies were inspired by the fake news warnings commonly seen on platforms like Facebook (for a review, refer to Kemp et al., 2024). However, it is crucial to recognize that, in the real world, the mere presence of a fact-checking flag does not guarantee user engagement, as individuals may choose not to click on it and see the correction that follows. To the best of my knowledge, no studies have investigated how fact-checking flags and reminder-based corrections influence actual click-through rates. Investigating this aspect is crucial because it could be the case that when people are re-exposed to fake news but do not see the correction due to negligence in clicking to view it, there might be an increased risk of a familiarity-driven misattribution error. This is because repeated exposure to fake news with reminders and increasing the information's salience through fact-checking flags may increase the believability of the information and the likelihood of being automatically activated at retrieval.

Thirdly, the headline stimuli employed in this research program were consistent or highly similar across the three papers. In the first empirical paper, headlines were sourced from fact-checking websites such as Snopes and PolitiFact, along with myths. Subsequently, in the second and third empirical papers, headlines exclusively from fact-checking websites were employed, but I introduced related images to replicate the appearance of news headlines on Google search engines, thereby enhancing their realistic portrayal. An advantage of using a similar material set is that it facilitates direct comparisons across studies, which is crucial for building a cumulative body of knowledge on effective fake news correction strategies. Furthermore, the observed consistency in outcomes across studies using the same material set underscores the reliability and reproducibility of the effects, contributing to the overall robustness of the findings.

However, a limitation of using the same stimuli set is that the findings may not generalize to real-world topics beyond those covered here. As stated above the stimuli set was carefully curated to include a diverse set of topics (i.e., health, sports, politics, crime) by extracting headlines from fact-checking websites. But, despite our efforts to include a diverse range of topics, the dynamic nature of misinformation in the real world means that some topics were not represented in our stimuli set. Acknowledging this, future research should consider broadening the scope of stimulus materials beyond fact-checking websites to capture a more comprehensive representation of current events. Collectively, studies using a wider variety of everyday and recent headline stimuli would provide more generalizable outcomes.

In a similar vein, future research should investigate how various correction strategies impact information when accompanied by artificial intelligence (AI)-generated images. The images in this research program were chosen based on their subjective relevance to headline details and their non-probative nature (i.e., they did not convey information about the veracity of the headline). Evidence indicates that statements accompanied by non-probative images can enhance the processing fluency of a statement, consequently increasing its perceived believability (for a review, see Zhang et al., 2021). It is plausible to assume that AI-generated images are more semantically related to the headline content than the stimuli images selected in the current research program as they are created using sophisticated algorithms that are often based on semantic understanding (i.e., ChatGPT). Consequently, this high semantic relatedness may enhance the processing fluency of the information, in turn increasing its perceived accuracy. Though speculative, empirical investigation on this is necessary, given the escalating prevalence of AI-generated content today.

Fourth, in the current research program, I recruited and tested participants using various methods. In the first paper, I tested UNCG undergraduate students via Zoom; in the second paper, I tested UNCG undergraduate students in the laboratory, and in the third paper, I recruited and tested participants through an open online platform (i.e., Prolific). Despite differences in testing methods and experimental paradigms, the consistency in patterns of conditional analyses for younger adults across the papers is noteworthy. This consistency is particularly valuable considering the varying degrees of experimenter supervision resulting from testing participants online or in person. It suggests that our findings may apply to real-world contexts where experimenter supervision is not evident. However, despite this, I acknowledge that our studies focused on the cultural context of American misinformation, and our sample exclusively comprised U.S. citizens. Consequently, generalizing the findings to the broader younger adult population may be limited. While lacking a theoretical rationale to anticipate that the observed findings would be moderated by culture or society, I am unable to empirically test this possibility. Future research endeavors should delve into how these findings translate to other cultural contexts and countries (cf. Carey et al., 2022).

Moving beyond the younger adult samples, the third empirical paper included older adults recruited from the same online platform. Incorporating older adults in the sample is important given that most prior studies have neglected this demographic (but see, Geraci & Guillory, 2013; Swire et al., 2017a; Kreps & Kriner, 2022), leaving age differences in effects of fake news corrections relatively understudied. However, it is noteworthy that research shows that samples from Prolific tend to be mostly white, well-educated, avid technology users (Turner et al., 2020), and more motivated to provide thoughtful responses to researchers' questions (Douglas et al., 2023). Consequently, this recruitment method raises concerns about the generalizability of our results to the broader population of older adults and underscores the importance of future work to include a diverse set of older adults who may show literacy regarding digital media and are more susceptible to internet-based fake news (Brashier & Schacter, 2020).

Conclusion

The primary goal of the three empirical papers presented in this dissertation was to explore how various approaches that may enhance the accessibility of fake news influence downstream memory and belief accuracy for news headlines. Our findings revealed that both reminders and the retrieval of fake news contributed to improved memory and belief accuracy for real news headlines, while repeated exposure to fake news did not improve nor impair memory or belief accuracy. Moreover, increasing the accessibility of fake news heightened the ability to detect corrections and integration of both the fake and real news headlines, and this was associated with higher memory accuracy for real news headlines when recollection-based retrieval was engaged. These findings add to the ongoing theoretical debate proposing that corrections should reiterate fake news details and highlight the importance of detection and recollection in opposing the unwanted effects of familiarity. From a practical lens, these findings suggest that correction strategies that are directed at enhancing integrative encoding and recollection-based retrieval may have promise to mitigate the effects of fake news. However, such correction strategies also run the risk of inducing familiarity backfire when recollection-based retrieval fails, showing how they are a double-edged sword. Moving forward, an interdisciplinary approach that incorporates both cognitive and social mechanisms, involving diverse samples (i.e., age, digital media literacy, education), should be considered in future research. This broader perspective would provide a more precise understanding of the effects of memory-based correction strategies and the intricate association between memory and belief accuracy.

REFERENCES

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Autry, K. S., & Duarte, S. E. (2021). Correcting the unknown: Negated corrections may increase belief in misinformation. *Applied Cognitive Psychology*, *35*(4), 960–975. https://doi.org/10.1002/acp.3823

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), 1–21. https://doi.org/10.3758/BF03209454

Bacon, F. T. (1979). Credibility of repeated statements: Memory for trivia. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(3), 241.

Bartsch, L. M., Loaiza, V. M., & Oberauer, K. (2019). Does limited working memory capacity underlie age differences in associative long-term memory? *Psychology and Aging*, *34*(2), 268–281. https://doi.org/10.1037/pag0000317

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

Bäuml, K.-H., Pastötter, B., & Hanslmayr, S. (2010). Binding and inhibition in episodic memory—Cognitive, emotional, and neural processes. *Neuroscience & Biobehavioral Reviews*, *34*(7), 1047–1054. https://doi.org/10.1016/j.neubiorev.2009.04.005

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*(4), 446–458. https://doi.org/10.1037/0096-3445.121.4.446

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

Bishara, A. J., & Jacoby, L. L. (2008). Aging, spaced retrieval, and inflexible memory performance. *Psychonomic Bulletin & Review*, *15*(1), 52–57. https://doi.org/10.3758/PBR.15.1.52

Bjork, R. A. (1970). Positive forgetting: The noninterference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior*, *9*(3), 255-268.

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review*, *110*(4), 762–784. https://doi.org/10.1037/0033-295X.110.4.762

Brashier, N. M., & Schacter, D. L. (2020). Aging in an era of fake news. *Current Directions in Psychological Science*, *29*(3), 316–323. https://doi.org/10.1177/0963721420915872

Brashier, N. M., Pennycook, G., Berinsky, A. J., & Rand, D. G. (2021). Timing matters when correcting fake news. *Proceedings of the National Academy of Sciences*, *118*(5), e2020043118. https://doi.org/10.1073/pnas.2020043118

Brashier, N. M., Umanath, S., Cabeza, R., & Marsh, E. J. (2017). Competing cues: Older adults rely on knowledge in the face of fluency. *Psychology and Aging*, *32*(4), 331–337. https://doi.org/10.1037/pag0000156

Brydges, T., Retamal, M., & Hanlon, M. (2020). Will COVID-19 support the transition to a more sustainable fashion industry? *Sustainability: Science, Practice and Policy*, *16*(1), 298–308. https://doi.org/10.1080/15487733.2020.1829848

Butler, L. H., Fay, N., & Ecker, U. K. H. (2023). Social endorsement influences the continued belief in corrected misinformation. *Journal of Applied Research in Memory and Cognition*, *12*(3), 364–375. https://doi.org/10.1037/mac0000080

Butler, L. H., Lamont, P., Wan, D. L. Y., Prike, T., Nasim, M., Walker, B., Fay, N., & Ecker, U. K. H. (2023). The (mis)information game: A social media simulator. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02153-x

Butterfuss, R., & Kendeou, P. (2020). Reducing interference from misconceptions: The role of inhibition in knowledge revision. *Journal of Educational Psychology*, *112*(4), 782–794. https://doi.org/10.1037/edu0000385

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*(3), 381-387. https://doi.org/10.1016/j.pec.2013.06.017

Cantarella, M., Fraccaroli, N., & Volpe, R. (2023). Does fake news affect voting behaviour? *Research Policy*, *52*(1), 104628. https://doi.org/10.1016/j.respol.2022.104628

Carey, J. M., Guess, A. M., Loewen, P. J., Merkley, E., Nyhan, B., Phillips, J. B., & Reifler, J. (2022). The ephemeral effects of fact-checks on COVID-19 misperceptions in the United States, Great Britain and Canada. *Nature Human Behaviour*, *6*(2), 236–243. https://doi.org/10.1038/s41562-021-01278-3

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology*, *66*(3), 460–473. https://doi.org/10.1037/0022-3514.66.3.460

Chan, M. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A Meta-Analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, *28*(11), 1531–1546. https://doi.org/10.1177/0956797617714579

Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., Kawata, A., Kovvuri, A., Martin, J., Morgan, E., Sandhu, M., Sang, R., Scholz-Bright, R., Welch, A. T., Wolff, A. G., Zhou, A., & Nyhan, B. (2020). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior*, *42*(4), 1073–1095. https://doi.org/10.1007/s11109-019-09533-0

Collier, J. R., Pillai, R. M., & Fazio, L. K. (2023). Multiple-choice quizzes improve memory for misinformation debunks, but do not reduce belief in misinformation. *Cognitive Research: Principles and Implications*, *8*(1), 37. https://doi.org/10.1186/s41235-023-00488-9

Connor Desai, S. A., Pilditch, T. D., & Madsen, J. K. (2020). The rational continued influence of misinformation. *Cognition*, 205, 104453. https://doi.org/10.1016/j.cognition.2020.104453

Connor Desai, S. A., & Reimers, S. (2019). Comparing the use of open and closed questions for Web-based measures of the continued-influence effect. *Behavior Research Methods*, *51*(3), 1426–1440. https://doi.org/10.3758/s13428-018-1066-zChan, M. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, *28*(11), 1531–1546. https://doi.org/10.1177/0956797617714579

Cook, J., Bedford, D., & Mandia, S. (2014). Raising climate literacy through addressing misinformation: Case studies in agnotology-based learning. *Journal of Geoscience Education*, *62*(3), 296–306. https://doi.org/10.5408/13-071.1

Danckert, S. L., & Craik, F. I. M. (2013). Does aging affect recall more than recognition memory? *Psychology and Aging*, *28*(4), 902–909. https://doi.org/10.1037/a0033263

De Saint Laurent, C., Murphy, G., Hegarty, K., & Greene, C. M. (2022). Measuring the effects of misinformation exposure and beliefs on behavioural intentions: A COVID-19 vaccination study. *Cognitive Research: Principles and Implications*, *7*(1), 87. https://doi.org/10.1186/s41235-022-00437-y

Ecker, U. K. H., & Ang, L. C. (2019). Political attitudes and the processing of misinformation corrections. *Political Psychology*, *40*(2), 241–260. https://doi.org/10.1111/pops.12494

Desai, S., & Reimers, S. (2019). Comparing the use of open and closed questions for web-based measures of the continued-influence effect. *Behavior Research Methods*, *51*(3), 1426–1440. https://doi.org/10.3758/s13428-018-1066-z

Ecker, U. K. H., Hogan, J. L., & Lewandowsky, S. (2017). Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition*, *6*(2), 185–192. https://doi.org/10.1016/j.jarmac.2017.01.014

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*(8), 1087–1100. https://doi.org/10.3758/MC.38.8.1087

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

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*(3), 570–578. https://doi.org/10.3758/s13423-011-0065-1

Ecker, U. K. H., O'Reilly, Z., Reid, J. S., & Chang, E. P. (2020). The effectiveness of short-format refutational fact-checks. *British Journal of Psychology*, *111*(1), 36–54. https://doi.org/10.1111/bjop.12383

Ecker, U. K., Lewandowsky, S., Chang, E. P., & Pillai, R. (2014). The effects of subtle misinformation in news headlines. *Journal of Experimental Psychology: Applied*, *20*(4), 323.

Ecker, U. K., Sharkey, C. X., & Swire-Thompson, B. (2023). Correcting vaccine misinformation: A failure to replicate familiarity or fear-driven backfire effects. *Plos one*, *18*(4), e0281140.

Evans, J., & Stanovich, K. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Fallis, D. (2015). What is disinformation?. *Library trends*, *63*(3), 401-426.

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, *144*(5), 993.

Fazio, L. K., Rand, D. G., & Pennycook, G. (2019). Repetition increases perceived truth equally for plausible and implausible statements. *Psychonomic Bulletin & Review*, *26*(5), 1705–1710. https://doi.org/10.3758/s13423-019-01651-4

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Fraundorf, S. H., Hourihan, K. L., Peters, R. A., & Benjamin, A. S. (2019). Aging and recognition memory: A meta-analysis. *Psychological Bulletin*, *145*(4), 339–371. https://doi.org/10.1037/bul0000185

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Freeman, D., Waite, F., Rosebrock, L., Petit, A., Causier, C., East, A., Jenner, L., Teale, A.-L., Carr, L., Mulhall, S., Bold, E., & Lambe, S. (2022). Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in England. *Psychological Medicine*, *52*(2), 251–263. https://doi.org/10.1017/S0033291720001890

Gabbert, F., Memon, A., & Wright, D. B. (2007). I saw it for longer than you: The relationship between perceived encoding duration and memory conformity. *Acta Psychologica*, *124*(3), 319–331. https://doi.org/10.1016/j.actpsy.2006.03.009

Galesic, M., Olsson, H., Dalege, J., van der Does, T., & Stein, D. L. (2021). Integrating social and cognitive aspects of belief dynamics: Towards a unifying framework. *Journal of The Royal Society Interface*, *18*(176), rsif.2020.0857, 20200857. https://doi.org/10.1098/rsif.2020.0857

Gallo, D. A. (2004). Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 120–128. https://doi.org/10.1037/0278-7393.30.1.120

Gordon, A., Quadflieg, S., Brooks, J. C. W., Ecker, U. K. H., & Lewandowsky, S. (2019). Keeping track of ‘alternative facts’: The neural correlates of processing misinformation corrections. *NeuroImage*, *193*, 46–56. https://doi.org/10.1016/j.neuroimage.2019.03.014

Grady, R. H., Ditto, P. H., & Loftus, E. F. (2021). Nevertheless, partisanship persisted: Fake news warnings help briefly, but bias returns with time. *Cognitive Research: Principles and Implications*, *6*(1), 52. https://doi.org/10.1186/s41235-021-00315-z

Greene, C. M., & Murphy, G. (2021). Quantifying the effects of fake news on behavior: Evidence from a study of COVID-19 misinformation. *Journal of Experimental Psychology: Applied*, *27*(4), 773–784. https://doi.org/10.1037/xap0000371

Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics* (Vol. 1). Wiley.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science*, *363*(6425), 374–378. https://doi.org/10.1126/science.aau2706

Guess, A., & Coppock, A. (2020). Does counter-attitudinal information cause backlash? Results from three large survey experiments. *British Journal of Political Science*, *50*(4), 1497–1515. https://doi.org/10.1017/S0007123418000327

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, *5*(1), eaau4586. https://doi.org/10.1126/sciadv.aau4586

Guillory, J. J., & Geraci, L. (2010). The persistence of inferences in memory for younger and older adults: Remembering facts and believing inferences. *Psychonomic Bulletin & Review*, *17*(1), 73–81. https://doi.org/10.3758/PBR.17.1.73

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*(4), 201–209. https://doi.org/10.1016/j.jarmac.2013.10.001

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, *16*(1), 107–112. https://doi.org/10.1016/S0022-5371(77)80012-1

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In Psychology of Learning and Motivation (Vol. 22, pp. 193–225). Elsevier. https://doi.org/10.1016/S0079-7421(08)60041-9

Hassan, A., & Barber, S. J. (2021). The effects of repetition frequency on the illusory truth effect. *Cognitive Research: Principles and Implications*, *6*(1), 38. https://doi.org/10.1186/s41235-021-00301-5

Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review*, *93*(3), 258.

Hay, J. F., & Jacoby, L. L. (1999). Separating habit and recollection in young and older adults: effects of elaborative processing and distinctiveness. Psychology and aging, 14(1), 122.

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, *50*(1), 264–284. https://doi.org/10.3758/s13428-017-0869-7

Hintzman, D. L. (2011). Research strategy in the study of memory: Fads, fallacies, and the search for the “coordinates of truth”. Perspectives on Psychological Science, 6(3), 253-271.

Höller, M. (2021). The human component in social media and fake news: The performance of UK opinion leaders on Twitter during the Brexit campaign. *European Journal of English Studies*, *25*(1), 80–95. https://doi.org/10.1080/13825577.2021.1918842

Inquisit 5. (2016). [Computer software]. https://www.millisecond.com

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541. https://doi.org/10.1016/0749-596X(91)90025-F

Jacoby, L. L. (1998). Invariance in automatic influences of memory: Toward a user's guide for the process-dissociation procedure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(1), 3–26. https://doi.org/10.1037/0278-7393.24.1.3

Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 3–22. https://doi.org/10.1037/0278-7393.25.1.3

Jacoby, L. L., Wahlheim, C. N., & Kelley, C. M. (2015). Memory consequences of looking back to notice change: Retroactive and proactive facilitation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(5), 1282–1297. https://doi.org/10.1037/xlm0000123

Jennings, J. M., & Jacoby, L. L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology and Aging*, *8*(2), 283–293. https://doi.org/10.1037/0882-7974.8.2.283

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*(6), 1420–1436. https://doi.org/10.1037/0278-7393.20.6.1420

Kahneman, D. (2011). *Thinking Fast and Slow*. Macmillan.

Kemp, P. L., Alexander, T. R., & Wahlheim, C. N. (2022b). Recalling fake news during real news corrections can impair or enhance memory updating: The role of recollection-based retrieval. *Cognitive Research: Principles and Implications*, *7*(1), 85. https://doi.org/10.1186/s41235-022-00434-1

Kemp, P. L., Goldman, A. C., & Wahlheim, C. N. (2024). On the role of memory in misinformation corrections: Repeated exposure, correction durability, and source credibility. *Current Opinion in Psychology*, *56*, 101783. https://doi.org/10.1016/j.copsyc.2023.101783

Kemp, P. L., Loaiza, V. M., & Wahlheim, C. N. (2022a). Fake news reminders and veracity labels differentially benefit memory and belief accuracy for news headlines. *Scientific Reports*, *12*(1), 21829. https://doi.org/10.1038/s41598-022-25649-6

Kemp, P., Wahlheim, C. N., & Loaiza, V. M. (2023). Correcting fake news headlines after repeated exposure: Memory and perceived accuracy in younger and older adults. Under review at *Cognitive Research: Principles and Implications*.

Kendeou, P., Butterfuss, R., Kim, J., & Van Boekel, M. (2019). Knowledge revision through the lenses of the three-pronged approach. *Memory & Cognition*, *47*(1), 33–46. https://doi.org/10.3758/s13421-018-0848-y

Kendeou, P., Walsh, E. K., Smith, E. R., & O'Brien, E. J. (2014). Knowledge revision processes in fefutation texts. *Discourse Processes*, *51*(5–6), 374–397. https://doi.org/10.1080/0163853X.2014.913961

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

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Zittrain, J. L. (2018). The science of fake news. *Science*, *359*(6380), 1094-1096.

Lenth, R. (2021). *emmeans: Estimated marginal means, aka least- squares means* (1.6.3) [Computer software]. https://cran.r-project.org/web/packages/emmeans/index.html

Levine, J. M., & Murphy, G. (1943). The learning and forgetting of controversial material. *The Journal of Abnormal and Social Psychology*, *38*(4), 507–517. https://doi.org/10.1037/h0062586

Lewandowsky, S., Cook, J., & Lombardi, D. (2020). *Debunking handbook 2020* [Data set]. Databrary. https://doi.org/10.17910/B7.1182

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful bebiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour*, *5*(3), 337–348. https://doi.org/10.1038/s41562-021-01056-1

Luna, K., Nogueira, M., & Albuquerque, P. B. (2019). Words in larger font are perceived as more important: Explaining the belief that font size affects memory. *Memory*, *27*(4), 555–560. https://doi.org/10.1080/09658211.2018.1529797

Lyons, B. A. (2023). Older Americans are more vulnerable to prior exposure effects in news evaluation. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-118

Maier, J., & Richter, T. (2013). Text belief consistency effects in the comprehension of multiple texts with conflicting information. *Cognition and Instruction*, *31*(2), 151–175. https://doi.org/10.1080/07370008.2013.769997

Makowski, D. (2021). *psycho: Efficient and publishing-oriented workflow for psychological science* (0.6.1) [Computer software]. https://cran.r-project.org/web/packages/psycho/index.html

Miller, M. L., Roe, D. J., Hu, C., & Bell, M. L. (2020). Power difference in a χ2 test vs generalized linear mixed model in the presence of missing data – a simulation study. *BMC Medical Research Methodology*, *20*(1), 50. https://doi.org/10.1186/s12874-020-00936-w

Moore, K. N., & Lampinen, J. M. (2016). The use of recollection rejection in the misinformation paradigm: Recollection rejection of misinformation. *Applied Cognitive Psychology*, *30*(6), 992–1004. https://doi.org/10.1002/acp.3291

Morrow, G., Swire-Thompson, B., Polny, J. M., Kopec, M., & Wihbey, J. P. (2022). The emerging science of content labeling: Contextualizing social media content moderation. *Journal of the Association for Information Science and Technology*, *73*(10), 1365–1386. https://doi.org/10.1002/asi.24637

Munnich, E., Ranney, M. A., & Song, M. (2007). Surprise, surprise: The role of surprising numerical feedback in belief change. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *29*(29), 503–508.

Newman, E. J., Swire-Thompson, B., & Ecker, U. K. H. (2022). Misinformation and the sins of memory: False-belief formation and limits on belief revision. *Journal of Applied Research in Memory and Cognition*, *11*(4), 471–477. https://doi.org/10.1037/mac0000090

Nyhan, B., Reifler, J., Richey, S., & Freed, G. L. (2014). Effective messages in vaccine promotion: A randomized trial. *Pediatrics*, *133*(4), e835–e842. https://doi.org/10.1542/peds.2013-2365

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging*, *23*(1), 104–118. https://doi.org/10.1037/0882-7974.23.1.104

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

Park, D. C., & Festini, S. B. (2017). Theories of memory and aging: A look at the past and a glimpse of the future. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *72*(1), 82–90. https://doi.org/10.1093/geronb/gbw066

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*(2), 299–320. https://doi.org/10.1037/0882-7974.17.2.299

Pehlivanoglu, D., Lighthall, N. R., Lin, T., Chi, K. J., Polk, R., Perez, E., Cahill, B. S., & Ebner, N. C. (2022). Aging in an "infodemic": The role of analytical reasoning, affect, and news consumption frequency on news veracity detection. *Journal of Experimental Psychology: Applied*, *28*(3), 468–485. https://doi.org/10.1037/xap0000426

Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science*, *66*(11), 4944–4957. https://doi.org/10.1287/mnsc.2019.3478

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, *147*(12), 1865–1880. https://doi.org/10.1037/xge0000465

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, *592*(7855), 590–595. https://doi.org/10.1038/s41586-021-03344-2

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, *80*, 34–72. https://doi.org/10.1016/j.cogpsych.2015.05.001

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science*, *31*(7), 770–780. https://doi.org/10.1177/0956797620939054

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39–50. https://doi.org/10.1016/j.cognition.2018.06.011

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, *25*(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

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. https://doi.org/10.1177/1075547015613523

Pluviano, S., Watt, C., & Della Sala, S. (2017). Misinformation lingers in memory: Failure of three pro-vaccination strategies. *PLOS ONE*, *12*(7), e0181640. https://doi.org/10.1371/journal.pone.0181640

Pluviano, S., Watt, C., Ragazzini, G., & Della Sala, S. (2019). Parents' beliefs in misinformation about vaccines are strengthened by pro-vaccine campaigns. *Cognitive Processing*, *20*(3), 325–331. https://doi.org/10.1007/s10339-019-00919-w

Pratkanis, A. R. (1989). The cognitive representation of attitudes. In R. Pratkanis, S. Breckler, & S. Greenwald (Eds.), *Attitude structure and function* (pp. 70–98). Erlbaum.

Prike, T., Blackley, P., Swire-Thompson, B., & Ecker, U. K. (2023). Examining the replicability of backfire effects after standalone corrections. *Cognitive Research: Principles and Implications*, *8*(1), 39.

Psychology Software Tools. (2012). *[E-Prime 2.0]* [Computer software]. https://www.pstnet.com

Psychology Software Tools. (2016). *[E-Prime 3.0]* [Computer software]. https://www.pstnet.com

Psychology Software Tools. (2020). *[E-Prime Go]* [Computer software]. https://www.pstnet.com

Putnam, A. L., Wahlheim, C. N., & Jacoby, L. L. (2014). Memory for flip-flopping: Detection and recollection of political contradictions. *Memory & Cognition*, *42*(7), 1198–1210. https://doi.org/10.3758/s13421-014-0419-9

R Core Team. (2021). *R: A language and environment for statistical computing. R Foundation for statistical computing*. https://www.R-project.org/.

R Core Team. (2022). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/.

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, *8*(3), 338-342.

Reinero, D. A., Harris, E. A., Rathje, S., Duke, A., & Van Bavel, J. J. (2023). Partisans are more likely to entrench their beliefs in misinformation when political outgroup members correct claims [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/z4df3

Reza, R., & Sunvy, A. S. (2023). The role of fact-checking sites during the Israel-Palestine conflict. *The Indonesian Journal of Communication Studies*, *16*(2), 122–134. https://doi.org/10.31315/ijcs.v16i2.11272

Rhodes, S., Greene, N. R., & Naveh-Benjamin, M. (2019). Age-related differences in recall and recognition: A meta-analysis. *Psychonomic Bulletin & Review*, *26*(5), 1529–1547. https://doi.org/10.3758/s13423-019-01649-y

Rich, P. R., & Zaragoza, M. S. (2016). The continued influence of implied and explicitly stated misinformation in news reports. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(1), 62–74. https://doi.org/10.1037/xlm0000155

Rich, P. R., & Zaragoza, M. S. (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. https://doi.org/10.1037/h0101850

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L. J., Recchia, G., van der Bles, A. M., & van der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society Open Science*, *7*(10), 201199. https://doi.org/10.1098/rsos.201199

Ryan, A. D., & Campbell, K. L. (2021). The ironic effect of older adults' increased task motivation: Implications for neurocognitive aging. *Psychonomic Bulletin & Review*, *28*(6), 1743–1754. https://doi.org/10.3758/s13423-021-01963-4

Sahakyan, L., & Foster, N. L. (2009). Intentional forgetting of actions: Comparison of list-method and item-method directed forgetting. *Journal of Memory and Language*, *61*(1), 134–152. https://doi.org/10.1016/j.jml.2009.02.006

Sahakyan, L., Delaney, P. F., Foster, N. L., & Abushanab, B. (2013). List-method directed forgetting in cognitive and clinical research. In *Psychology of Learning and Motivation* (Vol. 59, pp. 131–189). Elsevier. https://doi.org/10.1016/B978-0-12-407187-2.00004-6

Sanderson, J. A., Bowden, V., Swire-Thompson, B., Lewandowsky, S., & Ecker, U. K. H. (2023). Listening to misinformation while driving: Cognitive load and the effectiveness of (repeated) corrections. *Journal of Applied Research in Memory and Cognition*, *12*(3), 325–334. https://doi.org/10.1037/mac0000057

Sanderson, J. A., & Ecker, U. K. H. (2020). The challenge of misinformation and ways to reduce its impact. In P. Van Meter, A. List, D. Lombardi, & P. Kendeou (Eds.), *Handbook of Learning from Multiple Representations and Perspectives* (1st ed., pp. 461–476). Routledge. https://doi.org/10.4324/9780429443961-30

Santos, J., Sabino, A. L., Cruz, C. D., Mallari, J. K. C., Eusebio, M. J., & Gonzales, P. G. (2019). The effects of color coded words on memory retention. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3438889

Schacter, D. L. (2022). The seven sins of memory: An update. *Memory*, *30*(1), 37–42. https://doi.org/10.1080/09658211.2021.1873391

Schwarz, N., & Jalbert, M. (2020). When (fake) news feels true. In C. Mc Mahon (Ed.), *Psychological Insights for Understanding COVID-19 and Media and Technology* (1st ed., pp. 9–25). Routledge. https://doi.org/10.4324/9781003121756-2

Schwarz, N., Newman, E., & Leach, W. (2016). Making the truth stick & the myths fade: Lessons from cognitive psychology. *Behavioral Science & Policy; Washington*, *2*(1), 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. In *Advances in Experimental Social Psychology* (Vol. 39, pp. 127–161). Elsevier. https://doi.org/10.1016/S0065-2601(06)39003-X

Seifert, C. M. (2002). The continued influence of misinformation in memory: What makes a correction effective? In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 41, pp. 265–292). Academic Press.

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

Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research*, *31*(4), 713–724. https://doi.org/10.1086/426605

Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology and Aging*, *10*(4), 527–539. https://doi.org/10.1037/0882-7974.10.4.527

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*(2), 203–220. https://doi.org/10.3758/BF03196157

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 and Instruction*, *31*(2), 130–150. https://doi.org/10.1080/07370008.2013.769996

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin*, *111*(1), 42–61. https://doi.org/10.1037/0033-2909.111.1.42

Swire, B., Berinsky, A. J., Lewandowsky, S., & Ecker, U. K. H. (2017b). Processing political misinformation: Comprehending the Trump phenomenon. *Royal Society Open Science*, *4*(3), 160802. https://doi.org/10.1098/rsos.160802

Swire, B., Ecker, U. K. H., & Lewandowsky, S. (2017a). The role of familiarity in correcting inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(12), 1948–1961. https://doi.org/10.1037/xlm0000422

Swire-Thompson, B., Cook, J., Butler, L. H., Sanderson, J. A., Lewandowsky, S., & Ecker, U. K. H. (2021). Correction format has a limited role when debunking misinformation. *Cognitive Research: Principles and Implications*, *6*(1), 83. https://doi.org/10.1186/s41235-021-00346-6

Swire-Thompson, B., DeGutis, J., & Lazer, D. (2020). Searching for the backfire effect: Measurement and design considerations. *Journal of Applied Research in Memory and Cognition*, *9*(3), 286–299. https://doi.org/10.1016/j.jarmac.2020.06.006

Swire-Thompson, B., Dobbs, M., Thomas, A., & DeGutis, J. (2023). Memory failure predicts belief regression after the correction of misinformation. *Cognition*, *230*, 105276. https://doi.org/10.1016/j.cognition.2022.105276

Swire-Thompson, B., Miklaucic, N., Wihbey, J. P., Lazer, D., & DeGutis, J. (2022). The backfire effect after correcting misinformation is strongly associated with reliability. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001131

Tandoc Jr., E. C., Duffy, A., Jones-Jang, S. M., & Wen Pin, W. G. (2021). Poisoning the information well?: The impact of fake news on news media credibility. *Journal of Language and Politics*, *20*(5), 783–802. https://doi.org/10.1075/jlp.21029.tan

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*(1), 15.

Turner, A. M., Engelsma, T., Taylor, J. O., Sharma, R. K., & Demiris, G. (2020). Recruiting older adult participants through crowdsourcing platforms: Mechanical Turk versus Prolific Academic. *AMIA ... Annual Symposium Proceedings. AMIA Symposium*, *2020*, 1230–1238.

Udry, J., White, S. K., & Barber, S. J. (2022). The effects of repetition spacing on the illusory truth effect. *Cognition*, *225*, 105157. https://doi.org/10.1016/j.cognition.2022.105157

Unkelbach, C. (2007). Reversing the truth effect: learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 219.

Unkelbach, C., & Rom, S. C. (2017). A referential theory of the repetition-induced truth effect. *Cognition*, *160*, 110–126. https://doi.org/10.1016/j.cognition.2016.12.016

Unkelbach, C., Koch, A., Silva, R. R., & Garcia-Marques, T. (2019). Truth by repetition: Explanations and implications. *Current Directions in Psychological Science*, *28*(3), 247–253. https://doi.org/10.1177/0963721419827854

Vlasceanu, M., & Coman, A. (2021). The impact of social norms on health-related belief update. *Applied Psychology: Health and Well-Being*, aphw.12313. https://doi.org/10.1111/aphw.12313

Vlasceanu, M., Morais, M. J., & Coman, A. (2021). The effect of prediction error on belief update across the political spectrum. *Psychological Science*, *32*(6), 916–933. https://doi.org/10.1177/0956797621995208

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146-1151.

Vraga, E. K., & Bode, L. (2020). Defining misinformation and understanding its bounded nature: Using expertise and evidence for describing misinformation. *Political Communication, 37*(1), 136-144.

Wahlheim, C. N. (2014). Proactive effects of memory in young and older adults: The role of change recollection. *Memory & Cognition*, *42*(6), 950–964. https://doi.org/10.3758/s13421-014-0411-4

Wahlheim, C. N., Alexander, T. R., & Peske, C. D. (2020). Reminders of everyday misinformation statements can enhance memory for and beliefs in corrections of those statements in the short term. *Psychological Science*, *31*(10), 1325–1339. https://doi.org/10.1177/0956797620952797

Wahlheim, C. N., Garlitch, S. M., & Kemp, P. L. (2021). Context differentiation and remindings in episodic memory updating. *The Psychology of Learning and Motivation*, *75*, 245–277.

Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. *Memory & Cognition*, *41*(1), 1–15. https://doi.org/10.3758/s13421-012-0246-9

Wahlheim, C. N., Kemp, P., Wiley, R., & Engelhardt, A. (2023). Correction effects on the perceived accuracy of fake news headlines over time: The critical role of recollecting corrections. [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/tfdnu

Wahlheim, C. N., Smith, W. G., & Delaney, P. F. (2019). Reminders can enhance or impair episodic memory updating: A memory-for-change perspective. *Memory*, *27*(6), 849–867. https://doi.org/10.1080/09658211.2019.1582677

Wahlheim, C. N., & Zacks, J. M. (2019). Memory guides the processing of event changes for older and younger adults. *Journal of Experimental Psychology: General*, *148*(1), 30–50. https://doi.org/10.1037/xge0000458

Walter, N., & Tukachinsky, R. (2020). A meta-analytic examination of the continued Influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop It? *Communication Research*, *47*(2), 155–177. https://doi.org/10.1177/0093650219854600

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

Wilkes, A. L., & Leatherbarrow, M. (1988). Editing episodic memory following the identification of error. *The Quarterly Journal of Experimental Psychology Section A*, *40*(2), 361–387. https://doi.org/10.1080/02724988843000168

Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, *41*, 135-163.

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

APPENDIX A: SUPPPLEMENTAL MATERIAL FOR CHAPTER II

1. Statistical Methods

All analyses were conducted using R software (R Core Team, 2021). We examined the effects of interest using logistic and linear mixed-effects models from *lme4* (Bates et al., 2015). The models included fixed effects of Headline Type and Correction Classifications, where applicable. The models also included by-participant and by-item random intercepts. We performed Wald's χ^2 hypothesis tests using the *Anova* function of the *car* package (Fox & Weisberg, 2019) and post-hoc comparisons controlling for multiple comparisons using the Tukey method in the *emmeans* package (Lenth, 2021). The complete model specifications are available in the analysis scripts on the OSF: (https://osf.io/zg8yx/). The significance level was $\alpha = .05$.

2. Phase 1 Ratings

2.1 Familiarity Ratings in Phase 1

Table S1 (top row) displays the baseline familiarity ratings in Phase 1. In both experiments, participants perceived real news headlines as more familiar than fake news headlines in the reminder and unlabeled correction headlines, smallest *z* ratio = 2.82, *p* = .02. Familiarity ratings did not differ between real news headlines and fake news headlines associated with veracity labels in each experiment, largest *z* ratio = 1.38, *p* = .51. In both experiments, there were no significant differences in familiarity ratings across fake news headlines in the three correction conditions, largest *z* ratio = 1.86, *p* = .25.

2.2 Baseline Belief Ratings in Phase 1

Table S1 (bottom row) displays the baseline belief ratings in Phase 1. In both experiments, participants believed real news headlines more than fake news headlines in all three correction conditions, smallest *z* ratio = 5.25, *p* < .001. In Experiment 1, beliefs in fake news did not differ across the correction conditions, largest *z* ratio = 0.49, $p = .96$. In Experiment 2, including veracity-labels with fake news led to lower belief ratings than for all other conditions, smallest *z* ratio = 63.92, $p < .001$, showing that participants attended to the labels when making their ratings. Belief in fake news did not differ between the other two correction conditions, *z* ratio = 0.58, $p = .94$, replicating Experiment 1.

Table S1. Familiarity and Belief Ratings

	Experiment 1				Experiment 2			
Measure	Real News Repetition	Reminder + Correction	Labeled Correction	Unlabeled Correction	Real News Repetition	Reminder + Correction	Labeled Fake News	Unlabeled Correction
Familiarity	3.02 [2.77, 3.27]	2.85 [2.60, 3.10]	2.95 [2.70, 3.20]	2.87 [2.62, 3.12]	2.93 [2.70, 3.16]	2.76 [2.53, 2.99]	2.82 [2.59, 3.05]	2.74 [2.51, 2.97]
Baseline Beliefs	3.88 [3.69, 4.08]	3.53 [3.34, 3.73]	3.53 [3.33, 3.72]	3.51 [3.32, 3.70]	4.21 [4.06, 4.36]	4.00 [3.85, 4.15]	1.42 [1.28, 1.57]	3.98 [3.83, 4.12]

Note. Bootstrap 95% confidence intervals are displayed in brackets.

3. Multinomial Process Tree Analyses of Cued Recall

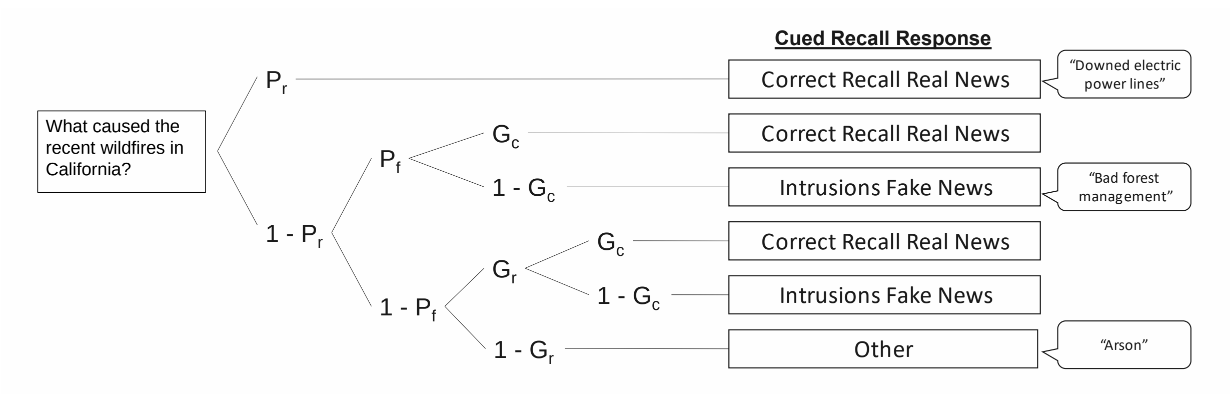
We fitted hierarchical Bayesian multinomial processing tree (MPT) models to the cued recall data for each experiment using the TreeBUGS package (Heck et al., 2018). MPT models are a class of measurement models that are used to estimate the probability of latent cognitive parameters from the frequency of observed categorical data. Here, the categorical data refer to the three possible cued recall response types: correct recall of real news from Phase 2, intrusions of fake news from Phase 1, and other types of intrusion errors. All ambiguous cued recall responses and responses to repeated real news headlines were not considered.

MPT models assume discrete cognitive states that lead participants to make one of these types of responses. Here, we used the independence model used extensively in similar research (e.g., Bartsch et al., 2018; Jacoby, 1999; Loaiza & Srokova, 2020). Figure S1 shows the model. This model assumes that recollection (*Pr*) can lead to correct recall of the real news headline. In the absence of recollection (1 – *Pr*), the fake news and real news headlines may both be familiar (*Pf*), in which case participants may guess randomly between headlines with equal probability, leading to either correct real news recall (*Gc* = 0.5) or intrusions of fake news (1 – *Gc*). Finally, in the absence of familiarity (1 – *Pf*), responses reflect random guessing with an equal probability between the correct response and incorrect fake news possibilities (*Gr* = 0.5) or an unrelated response (1 – *Gr*).

The hierarchical Bayesian MPT approach is advantageous because it accounts for heterogeneity between participants in the parameters while also affording a clear interpretation of null differences among conditions. The TreeBUGS package was used to fit the model with four Markov-Chain Monte-Carlo samples each comprising 100,000 iterations with 2,000 used for warm-up, 20,000 used for adaptation, and with a thinning factor of five. We confirmed good convergence between the chains both visually and by checking that the $\hat{R}$ statistic (i.e., the ratio of between-chain variance to within-chain variance) was close to one for all the parameters of the fitted models. We also ensured appropriate model fit by verifying that the *ppp* values were greater than .05.

To draw inferences about the effect of headline type on recollection and familiarity, we made within-subjects comparisons that computed the difference between the mean group posterior of one condition and that of another. Table S2 shows the posterior differences between the conditions. When CIs do not overlap with 0, we consider the difference credible. The analysis script to run or reproduce the model, as well as related information such as the model priors and its results, can be found on the OSF.

Figure S16. Cued Recall of Real and Fake News Details in Phase 3



Note. Independence multinomial model that estimates familiarity and recollection for responses in the cued recall task. R represents recollection of real news headlines and F represents familiarity of real news headlines. *Correct recall real news* are responses that included correct details from Phase 2 headlines. *Intrusions fake news* are responses that included details from false details from Phase 1 headlines. *Other* are responses that included details that were inconsistent with either correct or fake news headlines as well as omissions.

Table S2. Parameter Estimates of Recollection and Familiarity

Parameter	Experiment 1 Mean [95% CIs] Posterior Differences			Experiment 2 Mean [95% CIs] Posterior Differences		
	Reminder + Corrections vs. Labeled Corrections	Reminder + Corrections vs. Unlabeled Corrections	Labeled Corrections vs. Unlabeled Corrections	Reminder + Corrections vs. Labeled Fake News	Reminder + Corrections vs. Unlabeled Corrections	Labeled Fake News vs. Unlabeled Corrections
Recollection	**.11 [.05, .17]**	**0.26 [.19, .40]**	**.15 [.08, .22]**	**.20 [.12, .25]**	**.24 [.17, .31]**	.05 [-.02, .12]
Familiarity	.00 [-.05, .06]	**-.18 [-.27, -.09]**	**-.18 [-.27, -.09]**	-.01 [-.08, .06]	-.08 [-.18, .01]	-.07 [-.17, .02]

Note. CI = credibility interval. Bold values indicate credible differences in the posterior differences.

4. Power Analysis for Experiment 1

When planning Experiment 1, we used a standardized effect size estimate from the smallest effect of interest in Wahlheim et al. (2020) (Wahlheim et al., 2020) as a basis for a power analysis. That small-medium effect ($dz = .44$) corresponded to the finding that belief accuracy for misinformation intrusions was significantly greater when corrections had appeared with reminders rather than alone. At the time, our tool of choice for conducting power analyses was G*Power 3.1.9.2 (Faul et al., 2009). According to G*Power a sample size of 43 participants was sufficient to detect that effect size with 80% power (alpha = .05). However, our primary hypothesis about the effects of labeled corrections on overall memory and belief accuracy in the present study was that such corrections should lead to intermediate values relative to labeled corrections with fake news reminders and unlabeled corrections that appeared alone. We reasoned that those effect sizes may be smaller than the smallest effect size of interest from Wahlheim et al. (Wahlheim et al., 2020) We therefore chose to match the larger sample size from Wahlheim et al. ($N = 96$), which, according to G*Power, would allow us to detect a smaller effect size ($dz = .29$) with 80% power (alpha = .05). In the next section, a simulation analysis of Experiment 1 in the present study verifies that this sample size was more than sufficient to detect the smallest observed effect size of interest.

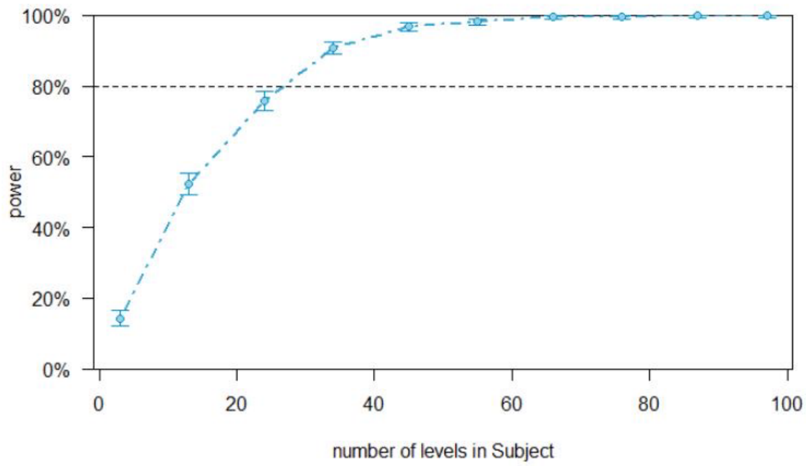
5. Sensitivity Analyses

5.1. Experiment 1 Sensitivity Analysis / Power Analysis for Experiment 2

We conducted a simulation-based sensitivity analysis using R software (R Core Team, 2021) based on the results from Experiment 1, which also served as a power analysis for planning the sample size in Experiment 2. The smallest effect size of interest in Experiment 1 corresponded to the difference in correct real news recall in Phase 3 between veracity-labeled corrections alone and those following fake news reminders. To examine the sensitivity to detect this effect, we first calculated the odds ratio of the pairwise difference between these conditions. To do this, we modeled the effects of experimental manipulations in Experiment 1 on real news recall in Phase 3 using a logistic mixed effects model, fitted with the *glmer* function from *lme4* (Bates et al., 2015). We included the Headline Type predictor as a fixed effect and included by-participant and by-item random intercepts. We then conducted a significance test (z test) to derive the log odds ratio effect size for the difference between the conditions of interest using the *dotest* function from *simr* (Green & MacLeod, 2016). We converted the log odds ratio to an odds ratio using the *exp* function from *base R* and interpreted the effect size using the *interpret_oddsratio* function from the *effectsize* package (Ben-Shachar et al., 2020). The odds ratio for effect size of interest was small (*OR* = 1.68 (Chen et al., 2010)).

We conducted a simulation-based sensitivity analysis using *simr* (Green & MacLeod, 2016) to examine the power to detect the odds ratio from the comparison described above. A sensitivity analysis based on 1,000 simulations with alpha set at .05 revealed that with 97 participants, Experiment 1 had 99.90% [*95% CI* = 99.44, 100.00] power to detect a small effect (*OR* = 1.68). To further determine the sample size necessary to detect this effect we generated power curve showing power levels across varying sample sizes. The power curve below (see Figure S2) shows that based on 1,000 simulations with 80% power (alpha = .05), a sample size of approximately 25 participants was sufficient required to detect an *OR* = 1.68. Experiment 1, which included 96 participants, was thus well powered for the smallest effect of interest.

Figure S2. Experiment 1 Power Curve

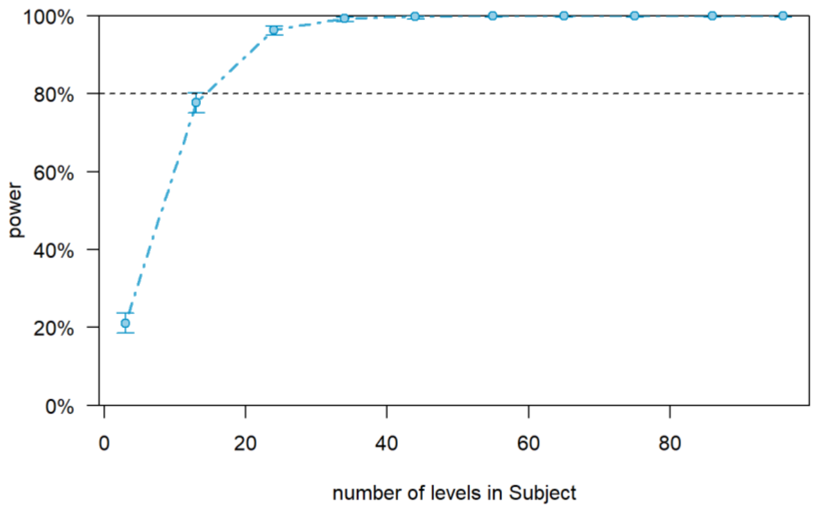


Note. Power curve to detect a very small effect (*OR* = 1.68) as a function of sample size (number of levels in Subject) using data from Experiment 1.

5.2. Experiment 2 Sensitivity Analyses

Similar to Experiment 1, to examine the sensitivity to detect the smallest effect size of interest, we conducted a simulation-based sensitivity analysis in the same manner as for Experiment 1. In Experiment 2, the smallest effect of interest was the difference in fake news recall in Phase 3 between for the conditions with veracity-labeled fake news in Phase 1 and fake news reminders before veracity-labeled headlines in Phase 2. This effect was of interest because it assessed differences in correction types on real news recall. A sensitivity analysis based on 1,000 simulations with alpha set at .05 revealed that with a sample size of 96 participants, Experiment 2 had 100% [*95% CI* = 99.63, 100.00] power to detect a small effect (*OR* = 2.12). A power curve based on 1,000 simulations with 80% power and alpha set at .05 (Figure S3) indicated that a sample size of 16 participants would have been sufficient to detect an effect of this size. Experiment 2 was therefore well-powered to detect the smallest effect of interest.

Figure S3. Experiment 2 Power Curve



Note. Power curve to detect a small effect (*OR* = 2.12) as a function of sample size (number of levels in Subject) using data from Experiment 2.

6. Attention Checks

We pre-registered a plan to exclude participants who failed more than two attention checks. However, the data collection rate was far slower than we anticipated given the virtual collection method. To meet our pre-registered sample size of at least 96 participants, we decided to include all participants in the analyses, regardless of their performance on the attention checks. For failing two attention checks, we would have excluded 11 participants in Experiment 1 and 10 participants in Experiment 2. Visual inspection of the data showed that although participants who we would have excluded sometimes performed poorly, many of them performed comparably to the other participants and vice versa. We therefore concluded that participants who failed more than two attention checks were unlikely to be categorically different from the others. To the extent that differences in task engagement was associated with memory and beliefs, we accounted for that by including the frequency of passed attention checks as a fixed effect (regressor of non-interest) in the mixed effects models. Moreover, reliable participant-level differences were captured in the by-participant random intercepts of the models.

7. Exploratory Analyses

In our pre-registrations, we outlined several exploratory questions motivated by previous findings (Pennycook et al., 2018; Wahlheim et al., 2021). However, upon further reflection, we decided to constrain our report to focus only on the exploratory questions that were most relevant to our overarching goals. Additional exploratory analyses that were not included in the manuscript or the present document can be found here: https://osf.io/zg8yx/. Below, we discuss findings from exploratory analyses concerning how 1) belief accuracy changes for fake news headlines in Phase 1 that eventually produced intrusions in Phase 3 (SI 7.1) and 2) whether patterns of cued recall performance remain constant when only considering responses that people believed were true (SI 7.2). We also tested whether analytic thinking that serves to correct faulty intuitions (Evans & Frankish, 2009; Stanovich, 2004) co-varies with memory and belief accuracy when recalling corrections of fake news using a variant of the cognitive reflection test (Frederick, 2005; Thomson & Oppenheimer, 2016) as an individual differences measure (SI 7.3).

7.1 Comparing Belief Ratings in Phases 1 and 3

Based on previous work showing that prior exposure to fake news increases perceived accuracy (Pennycook et al., 2018), it is possible that presenting fake news in Phase 1 and again in Phase 2 (along with the correction headline that repeats features of the fake news headline) results in greater overall belief in fake news at test relative to baseline beliefs in the current study. To assess this, we compared belief ratings during Phases 1 and 3 across all the correction conditions for instances when fake news intruded on the cued recall test (see Table S3). Experiment 1 showed that fake news was perceived as more believable in Phase 3 than in Phase

1, smallest *t* ratio = 3.97, $p < .001$. In Experiment 2, there was a significant interaction, χ^2 (2) = 159.58, $p < .001$, showing that fake news was perceived as more believable in Phase 3 than Phase 1 when such headlines were labeled as false in Phase 1, *t* ratio = 15.03, $p < .001$. There were no differences in belief ratings for fake news when corrections were unlabeled or fake news reminders preceded labeled corrections in Phase 2, largest *t* ratio = 1.34, $p = .18$.

Table S3. Belief Ratings for Headlines that Produced Intrusions of Fake News in Phases 1 and 3

	Experiment 1			Experiment 2		
Phase	Reminder + Corrections	Labeled Corrections	Unlabeled Corrections	Reminder + Corrections	Labeled Fake News	Unlabeled Corrections
Phase 1	3.90 [3.58, 4.22]	3.79 [3.49, 4.09]	3.86 [3.59, 4.13]	4.03 [3.77, 4.29]	1.66 [1.42, 1.90]	4.17 [3.96, 4.39]
Phase 3	4.17 [3.85, 4.49]	4.02 [3.72, 4.32]	4.38 [4.11, 4.65]	3.83 [3.57, 4.09]	3.62 [3.39, 3.86]	4.25 [4.04, 4.47]

Note: Bootstrap 95% confidence intervals are displayed in brackets.

7.2. Memory Accuracy for Responses Perceived as True

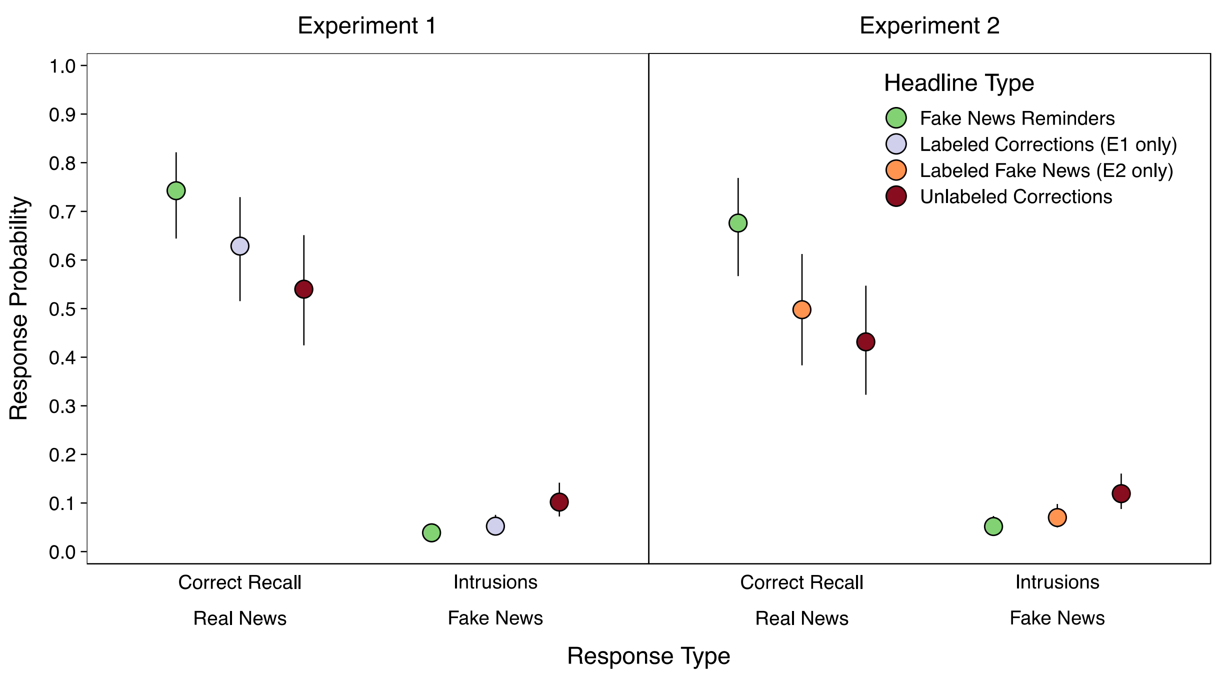
Correct Real News Recall

We examined whether the cued recall pattern changed when only considering responses perceived as true following across experiment comparisons manipulating report criterion based on believability (Wahlheim et al., 2020). A model was fitted to correct real news recall with Headline Type as a factor for corrected headlines that were perceived as true at test (see Figure S4). Parallel to the findings in the main text, Experiment 1 showed significantly higher real news recall for fake news reminder headlines than when corrections appeared with or without veracity-labels, smallest z ratio = 5.01, $p < .001$. Additionally, labeling corrections led to significantly higher real news recall than unlabeled corrections, z ratio = 3.55, p = .001. Experiment 2 showed significantly higher real news recall when fake news reminders immediately preceded corrections (M = .68 [*95% CI* = .57, .77]) regardless of whether fake news headlines appeared with veracity labels in Phase 1 before being corrected by unlabeled real news in Phase 2, smallest z ratio = 7.09, $p < .001$. In contrast to the main text, real news recall was higher when fake news appeared with a veracity label in Phase 1 (M = .50 [*95% CI* = .38, .61]) than unlabeled corrections headlines in Phase 2 (M = .43 [*95% CI* = .32, .55]), z ratio = 2.63, p = .02. This discrepancy indicated that guessing biases differed based on whether participants perceived their responses as true or false. More work is needed to understand this difference.

Intrusions of Fake News

A model was fitted to fake news intrusion errors with Headline Type as a factor for corrected headlines that were perceived as true at test (see Figure S4). Consistent with the main text, both experiments showed that corrections appearing with veracity labels, regardless of whether fake news reminders appeared in Phase 2 led to lower intrusion rates than presenting corrections without veracity labels (Experiment 1: *M* = .10 [*95% CI* = .07, .14]; Experiment 2: *M* = .12 [*95% CI* = .09, .16]), smallest *z* ratio = 4.61, *p* < .001. There were no significant differences in intrusions depending on whether fake news reminders were provided (Experiment 1: *M* = .04 [*95% CI* = .03, .06]; Experiment 2: *M* = .05 [*95% CI* = .04, .07]), corrections featured veracity-labels (*M* = .05 [*95% CI* = .04, .08]), or fake news appeared with veracity-labels in Phase 1 (*M* = .07 [*95% CI* = .05, .10]), smallest *z* ratio = 2.07, *p* = .10.

Figure S4. Cued Recall of Real and Fake News Details in Phase 3 as a Function of Perceived Truth



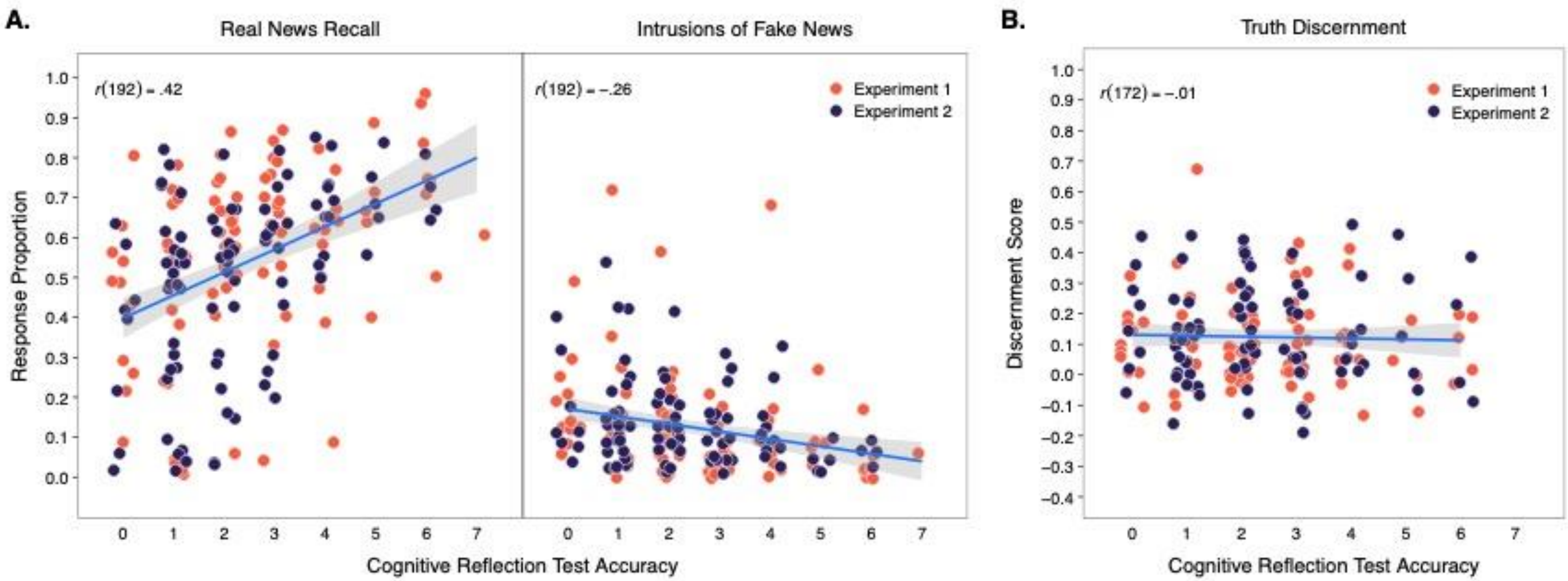
Note. Probabilities of real news recall and intrusions of fake news conditioned on perceived truth at test for each correction headline type condition. Points are probabilities estimated from mixed effects models; errors bars are 95% confidence intervals.

7.3. Cognitive Reflection and its Association with Memory and Belief Accuracy

Analytic thinking may be associated with memory and belief accuracy based on work reporting positive associations between the number of correct responses on cognitive reflection tests and discernment between real and fake news headlines (Pennycook & Rand, 2021). One possibility is that individual differences in recollection corresponds with the use of analytic thinking to overcome the pull of fake news familiarity during memory and belief judgments. This is consistent with the view from a dual-process theory that analytic thinking serves to correct faulty intuitions (Evans & Frankish, 2009; Stanovich, 2004). Using a seven-item cognitive reflection test, we examined how strongly cognitive reflection was associated with memory and belief accuracy. We correlated cognitive reflection, measured as the number of correct responses on the cognitive reflection test, with correct recall of real news, intrusions of fake news, and truth discernment for these responses computed as belief ratings for real news recall minus intrusions of fake news for each participant. Note that there were fewer observations for truth discernment because not all participants produced both responses at least once.

Figure S5 shows that cognitive reflection was positively associated with real news recall, $r(192) = .42$, $p < .001$, negatively associated with intrusions of fake news, $r(192) = -.26$, $p < .001$, and was not associated with truth discernment, $r(172) = -.01$, $p = .86$. These results are partially consistent with the view that analytic thinking serves to correct faulty information, but the absence of its association with truth discernment implies a more complex relationship among analytic thinking, memory, and beliefs. Differences in perceived accuracy that co-vary with analytic thinking may reflect memory differences upon which beliefs are sometimes based. However, more work is needed to understand the cognitive differences underlying truth discernment given concerns about whether the cognitive reflection test measures analytic thinking ability per se (Pennycook & Rand, 2019; Sinayev & Peters, 2015).

Figure S5. Cognitive Reflection and its Associations with Real News Recall, Intrusions of Fake News, and Truth Discernment in Beliefs in Cued Recall Responses



Note. The scatterplots display the association between cognitive reflection test accuracy and real news recall, intrusions of fake news, and truth discernment for each individual subject. The colored points represent individual participants from each experiment. Shaded regions are 95% confidence intervals around regression lines fitted to the data from both experiments.

Table S4. Proportions of Correction Classifications on the Cued Recall Test in Phase 3

	Experiment 1			Experiment 2		
Classification Type	Reminder + Corrections	Labeled Corrections	Unlabeled Corrections	Reminder + Corrections	Labeled Fake News	Unlabeled Corrections
Correction + Fake News Recalled	.58	.49	.44	.50	.40	.36
Correction + Fake News Not Recalled	.29	.29	.26	.30	.30	.26
Not a Correction + Fake News Not Recalled	.14	.22	.30	.20	.30	.39

Note. The values above correspond to the cell sizes of conditional real news recall and intrusions of fake news in Figure 3.

APPENDIX B: SUPPPLEMENTAL MATERIAL FOR CHAPTER III

8. Sensitivity Analyses

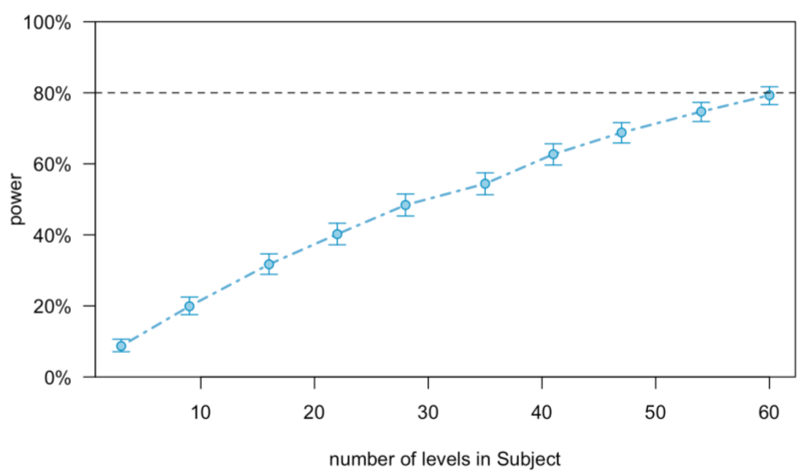
8.1. Experiment 1 Sensitivity Analyses

We conducted a simulation-based sensitivity analysis using R software (R Core Team, 2020) based on the results from Experiment 1. The smallest effect size of interest in Experiment 1 corresponded to the difference in correct real news recall in Phase 3 between the Control and Correction headlines. This effect was of interest because it indexes the proactive effects of fake news exposure on memory for real news. To examine the sensitivity to detect this effect size, we first calculated the odds ratio of the pairwise difference between these conditions. To do this, we modeled the effects of experimental manipulations in Experiment 1 on real news recall in Phase 3 using logistic mixed effects models, fitted with the *glmer* function from the *lme4* package (Bates et al., 2015). We treated the predictor variable Headline Type as a fixed effect, and Subjects and Items as random intercept effects. We then conducted a significance test (z test) to derive the log odds ratio effect size for the difference between the Control and Correction headlines using the *dotest* function from the *simr* package (Green & MacLeod, 2016). We then converted the log odds ratio to an odds ratio using the *exp* function from *base R*. Finally, we used the *interpret_oddsratio* function from the *effectsize* package (Ben-Shachar et al., 2020) to interpret the size of the effect. The odds ratios for the advantage in real news recall for Correction over Control headlines was very small (OR = 1.36; Chen, 2010).

We then conducted the simulation-based sensitivity analysis using *simr* (Green & MacLeod, 2016) to examine the power to detect the odds ratio from the comparison described above. We fitted the real news recall data in Phase 3 with a logistic mixed-effects model that included a fixed effect of Correction headline type. A sensitivity analysis based on 1,000 simulations with alpha set at .05 revealed that with 48 participants, Experiment 1 had 70.70% [*95% CI* = 67.57, 73.31] power to detect a very small effect (*OR* = 1.36).

As a further step to determine what sample size would be required to detect a very small effect (*OR* = 1.36) with 80% power, we expanded the power stimulation curve to include a larger sample size. We increased the sample to 60 subjects using the *extend* function from the *simr* package. For the power curve analysis, we fitted the real news recall data in Phase 3 with a logistic mixed-effects model, including Correction headline type as the reference level. A power curve (see Figure S6) based on 1,000 simulations with 80% power and alpha set at .05 indicated that a sample size of 60 participants is required to detect an effect with a very small odds ratio (*OR* = 1.36). Thus, a sample size of this or larger would be appropriate in the second experiment.

Figure S6. Experiment 1 Power Curve



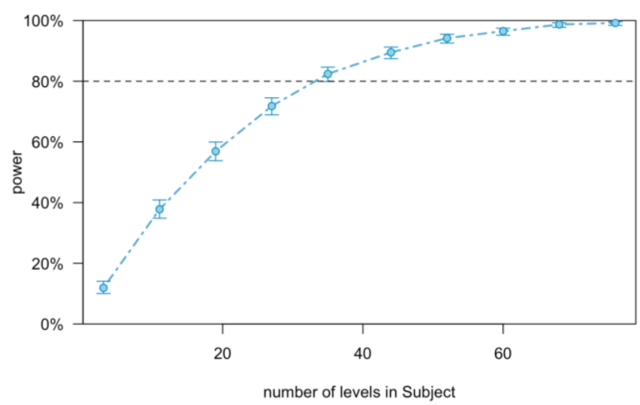
Note. Power curve to detect a very small effect (*OR* = 1.36) as a function of sample size (number of levels in Subject) using data from Experiment 1.

8.2. Experiment 2 Sensitivity Analyses

Similar to Experiment 1, to examine the sensitivity to detect the smallest effect size of interest, we conducted a simulation-based sensitivity analyses in the same manner as for

Experiment 1. In Experiment 2, the smallest effect of interest was the difference in correct real news recall in Phase 3 between the Control and Correction [Peers-Believe]) headlines. This effect was of interest because it assessed proactive effects of fake news exposure on memory for real news. Since was no difference in recall accuracy between the two Correction conditions, we selected the condition with the smallest difference from the Control condition. A sensitivity analysis based on 1,000 simulations with alpha set at .05 revealed that with a sample size of 76 participants, Experiment 2 had 99.50% [*95% CI* = 98.84, 99.84] power to detect a very small effect (*OR* = 1.55). A power curve based on 1,000 simulations with 80% power and alpha set at .05 (Figure S7) indicated that a sample size of only 37 participants would have been sufficient to detect a very small effect (*OR* = 1.55). Experiment 2 was therefore well-powered to detect the smallest effect of interest.

Figure S7. Experiment 2 Power Curve



Note. Power curve to detect a small effect (*OR* = 1.55) as a function of sample size (number of levels in Subject) using data from Experiment 2.

9. Cued Recall Scoring Method

Phase 3 cued recall responses were classified into one of four types. *Real news recall* responses included correct details from Phase 2 headlines. *Fake news intrusion error* included details from false details from Phase 1 headlines. *Ambiguous* responses did not differentiate between facts and misinformation. *Other errors* included details that were inconsistent with either correct or fake news headlines as well as omissions. Two raters who were blind to experimental conditions independently coded responses into these four categories after being on a set of responses from a pilot study. The initial interrater agreement was almost perfect for the Phase 3 real news recall response coding in Experiment 1, *Cohen's* $\kappa = .93$, $p < .001$. Note that a single rater coded the fake news recall responses during Phases 2 and 3. The initial interrater agreement was almost perfect Experiment 2, *Cohen's* $\kappa = .92$, $p < .001$, which included all responses in Phases 2 and 3. Discrepancies between raters were resolved through discussion.

10. Exploratory Analyses of Peer Belief Manipulation in Phase 1

10.1. Belief Ratings for Phase 1 Headlines

We first performed a manipulation check to determine whether participants on average believed more Phase 1 fake news headlines when most of their fictional peers also believed those headlines. For completeness, we also compared these with the rates of believing Phase 1 real news headlines. A model including Headline Type as a factor indicated a significant effect, $\chi^2(2) = 129.00$, $p < .001$, showing that participants indicated believing more fake news headlines in the Correction [Peers-Believe] ($M = .55$, *95% CI* = [.50, .60]) than Correction [Peers-Disbelieve] ($M = .33$, *95% CI* = [.29, .38]) condition, z ratio = 10.01, $p < .001$. Participants showed no significant difference in the proportion of headlines believed for real news in the Repetition condition ($M = .55$, *95% CI* = [.50, .60]) and fake news in the Correction [Peers-Believe] condition, z ratio = 0.16, $p = .99$, but they believed significantly more real news headlines in the Repetition condition than fake news headlines in the Correction [Peers-Disbelieve] condition, z

ratio = 9.86, $p < .001$. These results confirmed that perceived peer belief influenced participants' belief in fake news in the expected directions.

10.2. Belief Congruence, Correction Classification, and Fake News Recall

The cued recall results reported in the main manuscript showed enhanced memory updating associated with fake news recall in Phases 2 and 3. An exploratory aim of Experiment 2 was to determine if the measures of fake news recall in Phases 2 and 3 were sensitive to encoding differences in Phase 1 resulting from whether there was congruence between participant and peer beliefs. We tested the hypotheses that mismatches between participant and peer beliefs would stimulate more elaborative encoding and thus improve memory for fake news details in subsequent phases. Since fake news recall in Phases 2 and 3 was associated with facilitation in real news recall on the Phase 3 test, this also led to the hypothesis that real news recall would benefit from peer and participant belief mismatches. Performance on these memory measures was compared for *mismatched beliefs*, occurring when participants made a belief judgment that contradicted the fictional peer group, and *matched beliefs*, occurring when participants made a belief judgment that corresponded with the fictional peer group.

A Belief Congruence (Mismatched vs. Matched) × Phase (2 vs. 3) model fitted to fake news recall indicated a significant effect of Phase $\chi^2(1) = 30.68$, $p < .001$, showing that probabilities were significantly higher in Phase 2 ($M = .51$, *95% CI* = [.40, .63]) than in Phase 3 ($M = .37$, *95% CI* = [.27, .48]). Critically, there was no significant effect of Belief Congruence $\chi^2(1) = 0.47$, $p = .49$, showing that mismatches between participant and peer beliefs did not improve memory for fake news. There was no significant interaction, $\chi^2(1) = 0.04$, $p = .85$. Since belief congruence did not affect fake news recall in either phase, it was unlikely that it would affect real news recall in Phase 3. Indeed, a model fitted to real news recall in Phase 3 indicated no significant effect of Belief Congruence, $\chi^2(1) = 1.18$, $p = .28$.

Table S5. Observation Frequencies (Proportions) for Conditional Correct Recognition of Real News and False Recognition of Fake News: Experiment 1

		Detected				Not Detected			
		Younger		Older		Younger		Older	
Recognition	Conditional	Fake 1×	Fake 3×	Fake 1×	Fake 3×	Fake 1×	Fake 3×	Fake 1×	Fake 3×
Correct Real	Fake News Recognized	822 (.27)	909 (.30)	796 (.26)	909 (.30)	91 (.03)	98 (.03)	129 (.04)	96 (.03)
	Fake News Not Recognized	407 (.13)	351 (.11)	368 (.12)	328 (.11)	210 (.07)	172 (.06)	237 (.08)	197 (.06)
False Fake	Correction Remembered	975 (.32)	1043 (.34)	967 (.32)	1051 (.34)	149 (.05)	152 (.05)	213 (.07)	168 (.05)
	Correction Not Remembered	254 (.08)	217 (.07)	197 (.06)	186 (.06)	152 (.05)	118 (.04)	153 (.05)	125 (.04)

Note. Instances in which proportion totals do not sum 1 reflect rounding of individual proportions.

Table S6. Model Results for Correct Recognition of Real News as a Function of Fake News Exposure for Correction Headline Types Conditioned on Correction Classifications in Phases 2 and 3: Experiment 1

Effect	χ^2	df	p
Age Group	2.38	1	= .12
Correction Detection	77.65	1	< .001
Fake News Recognition	1272.84	1	< .001
Headline Type	6.84	1	< .01
Age Group × Correction Detection	1.57	1	= .21
Age Group × Fake News Recognition	0.41	1	= .52
Age Group × Headline Type	1.15	1	= .28
Correction Detection × Fake News Recognition	5.53	1	= .02
Correction Detection × Headline Type	0.73	1	= .39
Fake News Recognition × Headline Type	12.67	1	< .001
Age Group × Correction Detection × Fake News Recognition	4.65	1	= .03
Age Group × Correction Detection × Headline Type	0.22	1	= .64
Age Group × Fake News Recognition × Headline Type	4.50	1	= .03
Correction Detection × Fake News Recognition × Headline Type	2.15	1	= .14
Age Group × Correction Detection × Fake News Recognition × Headline Type	0.17	1	= .68

Table S7. Model Results for False Recognition of Fake News as a Function of Fake News Exposure for Correction Headline Types Conditioned on Correction Classifications in Phases 2 and 3: Experiment 1

Effect	χ^2	df	p
Age Group	0.35	1	= .55
Correction Detection	113.40	1	< .001
Correction Memory	544.06	1	< .001
Headline Type	7.74	1	< .01
Age Group × Correction Detection	< 0.01	1	= .96
Age Group × Correction Memory	0.25	1	= .62
Age Group × Headline Type	2.05	1	= .15
Correction Detection × Correction Memory	20.48	1	< .001
Correction Detection × Headline Type	0.05	1	= .82
Correction Memory × Headline Type	20.55	1	< .001
Age Group × Correction Detection × Correction Memory	0.38	1	= .54
Age Group × Correction Detection × Headline Type	1.96	1	= .16
Age Group × Correction Memory × Headline Type	0.86	1	= .35
Correction Detection × Correction Memory × Headline Type	1.50	1	= .22
Age Group × Correction Detection × Correction Memory × Headline Type	< 0.01	1	= .98

Table S8. Model Results for Perceived Accuracy in Phase 3: Experiment 1

Effect	χ^2	*df*	*p*
Age Group	2.32	1	= .13
Headline Type	2.85	1	= .09
Response Type	462.75	1	< .001
Age Group × Headline Type	0.61	1	= .44
Age Group × Response Type	2.02	1	= .15
Headline Type × Response Type	0.44	1	= .51
Age Group × Headline Type × Response Type	1.77	1	= .18

Table S9. Model Results for Perceived Accuracy in Phase 3 for Correct Recognition of Real News Conditioned on Correction Classifications in Phase 3: Experiment 1

Response Type	Effect	χ^2	df	p
Correct Recognition of Real News	Age Group	1.41	1	= .24
	Headline Type	< .01	1	= .92
	Correction Classification	138.36	1	< .001
	Age Group × Headline Type	3.26	1	= .07
	Age Group × Correction Classification	4.32	1	= .04
	Headline Type × Correction Classification	0.93	1	= .33
	Age Group × Headline Type × Correction Classification	2.66	1	= .10
False Recognition of Fake News	Age Group	2.58	1	= .11
	Headline Type	0.48	1	= .49
	Correction Classification	6.77	1	< .001
	Age Group × Headline Type	0.07	1	= .80
	Age Group × Correction Classification	0.23	1	= .63
	Headline Type × Correction Classification	5.42	1	= .02
	Age Group × Headline Type × Correction Classification	0.19	1	= .66

Table S10. Observation Frequencies (Proportions) for Conditional Correct Recall of Real News and False Recall of Fake News: Experiment 2

		Detected				Not Detected			
		Younger		Older		Younger		Older	
Recall	Conditional	Fake 1×	Fake 3×	Fake 1×	Fake 3×	Fake 1×	Fake 3×	Fake 1×	Fake 3×
Correct Real	Fake News Recalled	543 (.18)	670 (.22)	544 (.18)	664 (.22)	51 (.02)	42 (.01)	56 (.02)	62 (.02)
	Fake News Not Recalled	628 (.21)	526 (.17)	681 (.22)	585 (.19)	308 (.10)	292 (.10)	249 (.08)	219 (.07)
False Fake	Correction Remembered	840 (.27)	920 (.30)	964 (.32)	1028 (.34)	113 (.04)	98 (.03)	121 (.04)	115 (.04)
	Correction Not Remembered	331 (.11)	276 (.09)	261 (.09)	221 (.07)	246 (.08)	236 (.08)	184 (.06)	166 (.05)

Note. Instances in which proportion totals do not sum 1 reflect rounding of individual proportion.

Table S11. Model Results for Correct Recall of Real News as a Function of Fake News Exposure for Correction Headline Types Conditioned on Correction Classifications in Phases 2 and 3: Experiment 2

Effect	χ^2	df	p
Age Group	4.47	1	= .03
Correction Detection	95.43	1	< .001
Fake News Recall	656.24	1	< .001
Headline Type	13.29	1	< .001
Age Group × Correction Detection	0.32	1	= .57
Age Group × Fake News Recall	12.23	1	< .001
Age Group × Headline Type	0.22	1	= .64
Correction Detection × Fake News Recall	6.94	1	< .01
Correction Detection × Headline Type	0.59	1	= .44
Fake News Recall × Headline Type	1.70	1	= .19
Age Group × Correction Detection × Fake News Recall	2.63	1	= .10
Age Group × Correction Detection × Headline Type	0.65	1	= .42
Age Group × Fake News Recall × Headline Type	0.53	1	= .47
Correction Detection × Fake News Recall × Headline Type	0.85	1	= .36
Age Group × Correction Detection × Fake News Recall × Headline Type	0.07	1	= .79

Table S12. Model Results for Intrusions of Fake News as a Function of Fake News Exposure for Correction Headline Types Conditioned on Correction Classifications in Phases 2 and 3: Experiment 2

Effect	χ^2	df	p
Age Group	8.53	1	< .01
Correction Detection	62.41	1	< .001
Correction Memory	247.62	1	< .001
Headline Type	35.80	1	< .001
Age Group × Correction Detection	3.76	1	= .05
Age Group × Correction Memory	1.70	1	= .19
Age Group × Headline Type	0.59	1	= .44
Correction Detection × Correction Memory	36.18	1	< .001
Correction Detection × Headline Type	0.52	1	= .47
Correction Memory × Headline Type	7.82	1	< .01
Age Group × Correction Detection × Correction Memory	7.25	1	< .01
Age Group × Correction Detection × Headline Type	0.16	1	= .69
Age Group × Correction Memory × Headline Type	0.69	1	= .41
Correction Detection × Correction Memory × Headline Type	1.07	1	= .30
Age Group × Correction Detection × Correction Memory × Headline Type	0.11	1	= .75

Table S13. Model Results for Perceived Accuracy in Phase 3: Experiment 2

Effect	χ^2	df	p
Age Group	2.59	1	= .11
Headline Type	3.52	1	= .06
Correction Classification	252.25	1	< .001
Age Group × Headline Type	0.19	1	= .66
Age Group × Correction Classification	3.55	1	= .06
Headline Type × Correction Classification	0.52	1	= .47
Age Group × Headline Type × Correction Classification	< .01	1	= .95

Table S14. Model Results for Perceived Accuracy in Phase 3 for Intrusions of Fake News Conditioned on Correction Classifications in Phase 3: Experiment 2

Response Type	Effect	χ^2	df	p
Correct Recall of Real News	Age Group	4.98	1	= .03
	Headline Type	0.66	1	= .42
	Correction Classification	34.56	1	< .001
	Age Group × Headline Type	0.82	1	= .37
	Age Group × Correction Classification	< .01	1	= .99
	Headline Type × Correction Classification	0.04	1	= .84
	Age Group × Headline Type × Correction Classification	< .01	1	= .95
Intrusions of Fake News	Age Group	0.38	1	= .54
	Headline Type	1.24	1	= .27
	Correction Classification	10.60	1	= .001
	Age Group × Headline Type	0.03	1	= .85
	Age Group × Correction Classification	19.76	1	< .001
	Headline Type × Correction Classification	0.11	1	= .74
	Age Group × Headline Type × Correction Classification	0.72	1	= .40