UNIVERSITY OF CALIFORNIA

Santa Barbara

The Effects of Message, Source, and Audience Features on the Effectiveness of Misinformation Correction

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Communication

by

Xingyu Liu

Committee in charge:

Professor Miriam J. Metzger, Chair

Professor Andrew J. Flanagin

Professor Kristy A. Hamilton

December 2024

The dissertation of Xingyu Liu is approved.

_____

Andrew J. Flanagin

_____

Kristy A. Hamilton

_____

Miriam J. Metzger, Committee Chair

December 2024

The Effects of Message, Source, and Audience Features on the Effectiveness of Misinformation Correction

ACKNOWLEDGEMENTS

Words cannot express my gratitude to my advisor and chair of my committee for her invaluable patience and feedback. I couldn't have asked for a better advisor than Dr. Miriam Metzger, whose knowledge and experience guided me every step of the way, whose care and encouragement uplifted me in moments of challenge, and whose support carried me throughout this journey like a ray of sunshine.

I'm also grateful to my partner, Anthony, for always being there for me when I'm stressful or frustrated. Your belief in me have kept my spirits and motivation high during this process. And I would like to thank my family and friends for their emotional support, which transcended the distances of time and the Pacific Ocean.

Thank you and I would have not done this without all of you!

VITA OF XINGYU LIU
December 2024

EDUCATION

Bachelor of Arts in English, Sichuan University, June 2018
Master of Science in Psychology, Arizona State University, June 2020
Doctor of Philosophy in Communication, University of California, Santa Barbara, December 2024 (expected)

PROFESSIONAL EMPLOYMENT

2018-2019: Teaching Assistant, Department of Psychology, Arizona State University
2019-2020: Statistical Consultant, Department of Psychology, Arizona State University
2020-2024: Teaching Assistant, Department of Communication, University of California, Santa Barbara

PUBLICATIONS

"Selective Exposure and Credibility Perceptions of News on Social Media," Unpublished thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Psychology, Arizona State University, 2020. 61 pp.

"Checking the Fact-Checkers: The Role of Source Type, Perceived Credibility, and Individual Differences in Fact-Checking Effectiveness," *Communication Research*, 2023, https://doi.org/10.1177/00936502231206419

FIELDS OF STUDY

Major Field: News Credibility and Perception

Studies in Credibility Research with Professor Miriam Metzger

Studies in Cognition Research with Professor Kristy Hamilton

Studies in Research Methods and Statistics with Professors Miriam Metzger and Kristy Hamilton

ABSTRACT

The Effects of Message, Source, and Audience Features on the Effectiveness of Misinformation Correction

by

Xingyu Liu

The proliferation of misinformation has profound impacts on public discourse and decision-making. This dissertation explores the effectiveness of misinformation corrections, focusing on how message, source, and audience features of corrections influence responses. Specifically, with perceived language intensity of the correction as a central message feature, the study examines its unique and interactive effects with other factors in influencing correction effectiveness, including source credibility as a source feature, and attitude discrepancy and issue involvement as audience features. Drawing on the language expectancy theory (LET) and the information processing model, the Language Intensity and Correction Effectiveness Model (LICE) is proposed and examined under different contexts—for familiar and unfamiliar sources of corrections—where source credibility plays distinct roles in the processes.

To test the model, an experimental design was implemented in which participants were exposed to pro-attitudinal misinformation followed by corrections with varying levels of language intensity. These corrections were delivered by sources of differing familiarity

and credibility. Audience features such as prior beliefs in the veracity of the misinformation and issue involvement were measured. Correction effectiveness was operationalized as reduction in belief in the misinformation, by comparing perceived credibility of the misinformation after exposure to the correction to the control group with no correction, and between experimental conditions with varying levels of influencing factors. Message and source credibility perceptions of the correction, perceived language intensity of the correction, and demographics were also measured.

Findings suggest that corrections are generally effective in reducing belief in the misinformation regardless of language intensity, but high-intensity language has the potential to undermine correction effectiveness across varying levels of source credibility, attitude discrepancy, and issue involvement. Specifically, intense language negatively impacts correction effectiveness through decreased message credibility of the correction, which in turn decreases source credibility for unfamiliar sources. The detrimental effect of high-intensity language remains even for familiar sources with high prior credibility perceptions, and is particularly pronounced among individuals with low attitude discrepancy with the correction and high involvement with the issue.

Overall, this research reveals a distinct process in responses to misinformation corrections, where individuals critically examine all possible cues to assess credibility, emphasizing a balanced approach to consider the combined effects of message, source, and audience features to develop more targeted and effective correction messages for different source and audience groups.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction to Misinformation and its Correction

In a media era dominated by information overload and rapid dissemination, the problem of misinformation underscores the importance of robust fact-checking attempts and other mechanisms to safeguard the truth. Misinformation can be defined as false information that is spread either by mistake or with intent to mislead, although it is often called "disinformation" when there is intent to deceive (Lewandowsky et al., 2020). The origins of misinformation vary in terms of context and intent, including rumors and fiction, governments and politicians, corporate interests, and the media, with the Internet as a significant contributor to the spread of misinformation (Lewandowsky et al., 2012). Regardless of the source, misinformation has the potential to harm both individuals and society by, for example, imperiling public health or increasing partisan animosity. As an effort to combat misinformation, several strategies have emerged, such as prebunking, accuracy nudges, and debunking. Of these, debunking (i.e., misinformation correction) has been used most extensively, especially in the digital environment.

Debunking is different from other intervention strategies, including prebunking and accuracy nudges. *Prebunking* using psychological inoculation aims to reduce people's susceptibility to misinformation by pre-emptively exposing them to a weakened dose of the unwanted persuasion attempt, including a forewarning of an impending attack on one's beliefs or attitudes, and a refutation of the upcoming argument (Roozenbeek et al., 2023). For example, people are provided with a video counterarguing false claims on climate change or conferring general strategies commonly employed by those who spread misinformation (e.g., logical fallacies, emotional manipulation) *before* encountering the misinformation itself as a type of inoculation. Playing inoculation games that encourage people to generate their own

counterarguments to misinformation have also proved to be useful in some contexts (e.g., Basol et al., 2021; Roozenbeek & van der Linden, 2020; Saleh et al., 2021). However, it is not always possible to predict what misinformation people will encounter, and prebunking is not generally effective when the inoculation treatment is not specific enough in treating the exact misinformation content that people later encounter (Roozenbeek et al., 2023). The highly complex and interactive nature of social media also hamper prebunking's efficacy in preventing people from misinformation attacks, emphasizing the need for other techniques that are more effective in online environments.

*Accuracy nudge*s are specifically designed to shift people' attention toward accuracy and reduce their intention to share misinformation on social media. For example, in a typical accuracy nudge, people are asked to pause for a few seconds to consider the accuracy of news before deciding to share it, which has been shown to significantly reduce their willingness to share misinformation (Fazio, 2020). Accuracy primes that ask people to evaluate whether a headline is accurate also improve the quality of their subsequent news sharing decisions (Pennycook & Rand, 2022). Nevertheless, some replications failed to find effects of accuracy nudges on sharing decisions (e.g., Roozenbeek et al., 2021) and the effects of accuracy nudges in real-world social media environments are understudied. Therefore, although accuracy nudges are relatively easy to implement and cost-efficient on social media, their effectiveness is still under doubt. As such, debunking is a more popular approach to countering misinformation across news contexts and media platforms.

Debunking refers to correcting misinformation after it has been spread. It is nearly equivalent to fact-checking although a fact-check can rate a story to be true, whereas debunking only tackles false information. By correcting the fact, claim, logic, or

manipulation techniques that the misinformation uses (e.g., Vraga et al., 2020), the debunking approach can be applied in multiple news contexts (e.g., health, politics, science, entertainment) and presented on various communication platforms (e.g., social media, news website, traditional news sources). The debunking content may originate from various sources, including human experts and professionals (i.e., experienced journalists or staff trained in information verification methods), collective wisdom (i.e., community fact-checking through crowdsourcing), or it can be computer-generated using algorithms. Research shows that debunking or corrections are generally effective in reducing beliefs in misinformation. Meta-analyses reveal small to medium effective sizes for the effectiveness of misinformation corrections in changing beliefs or attitudes, which varies with a number of influencing factors and in different news contexts (e.g., Walter et al., 2020; Walter et al., 2019). This dissertation will focus on debunking because it is the predominant strategy used for countering misinformation. It will address obstacles to effective misinformation corrections, and will explore potential solutions (Chapter 1). It will also examine research on features of corrections that influence their effectiveness (Chapter 2) with the aim to fill gaps in the existing literature by proposing a comprehensive model of correction effectiveness centered on perceived language intensity of the correction and other factors (Chapters 3-6).

Obstacles to Effective Correction of Misinformation

Attempts to correct misinformation are not always successful. Researchers have been striving to identify and overcome the barriers that serve as obstacles to correction effectiveness to shed light on when corrections can be most effective. One of the most prominent obstacles is the continued influence effect (CIE), where people continue to rely on misinformation in their memory and reasoning even after a credible correction has been

presented (Johnson & Seifert, 1994). The CIE was first proposed by Johnson and Seifert (1994), and has been studied in the persuasion literature by both psychology and communication scholars (Ecker et al., 2014; Ecker et al., 2022; Lewandowsky et al., 2020).

There are currently two dominant cognitive explanations for the CIE. The first concerns the *mental-model* account, which suggests that failures of information integration and updating drive the CIE. This account assumes that people build mental models of unfolding events to reason and draw inferences from (Johnson & Seifert, 1994). Corrections that invalidate a piece of critical information in the unfolding event will create a gap in the mental model. Therefore, when asked about the event, people may fail to update and revise their memory, unless a plausible alternative is provided to fill in the gap. As a result, such failure to integrate new information into the model makes people more willing to discount the correction and still rely on the misinformation. The second explanation suggests that the CIE occurs due to a *failure at retrieval*. According to this view, misinformation and its correction compete for activation during retrieval through dual processing routes—automatic vs. strategic retrieval processes (Ayers & Reder, 1998). While misinformation is more likely to be activated by cues and automatically retrieved, corrections require a strategic assessment of the information validity to be retrieved. As the latter is more cognitively demanding, it often fails, which results in the continued retrieval of the misinformation.

Research has shown that the CIE can only be reduced but cannot be fully eliminated. In the classic laboratory paradigm, a correction is typically found to only halve the number of references to misinformation in responses, compared to a no-correction control condition, even when people can clearly remember and report the correction (Ecker et al., 2014; Ecker et al., 2017). Several backfire effects also appear to strengthen the CIE, such as the

familiarity backfire effect, overkill backfire effect, and worldview backfire effect. Each of these is explained below.

A backfire or "boomerang" effect of misinformation correction happens when the correction ironically increases belief or reliance on the misinformation relative to pre-correction or to a no-correction baseline (Lewandowsky et al., 2020). The *familiarity backfire effect* explains a situation when repetition of misinformation increases its familiarity, and corrections that unavoidably repeat misinformation produce the illusory truth effect (Schwarz et al., 2007) such that increased familiarity increases belief in or perceived accuracy of the misinformation (Autry & Duarte, 2021; Lewandowsky et al., 2020). For example, Berinsky (2017) found that merely repeating a rumor increased its believability by increasing fluency (i.e., the ease of information processing). Simply asking participants to rehearse the rumor without any information about its veracity increased their willingness to believe the rumor, even weeks after they initially read the rumor. And this influence holds even when the rumor was repeated with a strong correction. However, more recent research found no evidence of such a familiarity backfire effect, and instead found corrections explicitly repeating the misinformation are more effective in reducing reliance on misinformation than corrections that avoid repetition (e.g., Ecker et al., 2017; Ecker et al., 2023). This is probably because corrections make both the falsity of misinformation and the discrepancy between the truth and untruth more salient, which increases people's awareness and criticism of the misinformation.

The *overkill backfire effect* refers to the idea that too many corrections may produce unintended effects or even backfire (Lewandowsky et al., 2020). In the context of persuasion, too many persuasive arguments may lead to skepticism and reactance (Shu & Carlson, 2014),

and processing many arguments can also burden the cognitive load that reduces the persuasive effect (Schwarz et al., 2007). Therefore, studies have suggested that more corrections are not always better, and that less is more (Denner et al., 2023; Lewandowsky et al., 2012). For instance, in the experiment by Denner and colleagues (2023), participants first saw a Facebook post containing misinformation about a large producer of organic goods delivering fruits with harmful chemicals, and then read fictious posts of a food blog containing either no correction of the misinformation, one correction, or six corrections. Results showed that multiple corrections aiming to prove the innocence of the targeted food producer instead decreased people's trust in the producer through increased perceived persuasive intent and reactance, while a single correction increased trust in the producer. Nevertheless, another study found no evidence for such an effect of multiple corrections, such that a larger number of relevant corrections or counterarguments were conducive to greater reduction of beliefs in misinformation (Ecker et al., 2019). This finding is in line with typical persuasion patterns (Petty & Cacioppo, 1984) and the Bayesian view of belief change (Cook & Lewandowsky, 2016) where more arguments or evidence leads to greater belief revision.

A correction that challenges one's worldview may increase belief in the misinformation, which is called the *worldview backfire effect* (Lewandowsky et al., 2020). An individual's worldview may include one's prior beliefs or attitudes on an issue, political partisanship, and cultural values. Abundant evidence shows that corrections congruent with one's worldview often result in stronger persuasive impacts than corrections incongruent with one's worldview (e.g., Banas et al., 2022; Walter et al., 2019; Walter & Tukachinsky, 2020). This can be explained by motivated reasoning and confirmation bias, where people

tend to actively pursue, interpret, and believe information that confirms their prior beliefs, while engaging in motivated processing and discounting counter-attitudinal information, even in the face of concrete evidence and mounting facts (Kunda, 1990; Taber & Lodge, 2006). As a result, worldview-incongruent corrections are typically less effective than congruent ones and, under certain conditions, can even backfire. Evidence of the worldview backfire effect has been demonstrated across the domains of climate change (Hart & Nisbet, 2012), vaccination (Nyhan & Reifler, 2015), and other controversial political issues (Nyhan & Reifler, 2010), where people view the counter-attitudinal corrections as threats to their worldview and thus untrustworthy, counterproductively leading to increased belief in the misinformation. However, here again recent studies have failed to obtain such a backfire effect and revealed similar belief reduction in misinformation across individuals with different worldviews (Haglin, 2017; Swire, Berinsky, et al., 2017; Swire-Thompson et al., 2020; Wood & Porter, 2019). It is likely that other confounding factors influence people's acceptance of the correction (e.g., issue involvement, source credibility) that makes the backfire effects elusive under some conditions, and corrections in a particular context may be more prone to backfire effects than in another context (e.g., corrections in the political context may elicit more reactance than in the health context).

In sum, although many cognitive obstacles including the CIE and backfire effects have been identified, their hindering effects on the effectiveness of misinformation corrections are not consistent across studies and contexts. This also highlights the complexity of cognitive processes individuals employ when encountering misinformation and corrections. Other obstacles arise from the shape of the contemporary media landscape, such as the formation of echo chambers and filter bubbles. For example, users on social media are

often only exposed to attitude-consistent viewpoints and connected to in-group members due to personalization algorithms and selective exposure tendency (Barberá et al., 2015; Pariser, 2011). Societal factors including the decline of social capital, social inequality, and increasing political polarization also present fundamental challenges to misinformation correction (Sanderson & Ecker, 2020; Van Bavel et al., 2021; van der Linden, 2022).

Recommendations to Overcome Obstacles of Corrections

Regardless of the obstacles to misinformation correction, various message designs and debunking strategies can be used properly to overcome those barriers and increase correction effectiveness. Recommendations from the literature include (1) repeat corrections; (2) provide plausible alternatives; (3) provide worldview-congruent or self-affirming corrections; (4) use graphical representations; (5) provide corrections immediately after the misinformation; (6) use combinations of strategies.

First, the success of corrections can be enhanced if they are repeated. While initial research found evidence for a familiarity backfire effect, wherein corrections repeating the misinformation increase belief in it, as mentioned earlier subsequent studies have failed to replicate these findings (Ecker et al., 2011; Ecker et al., 2017; Ecker et al., 2023). Alternatively, any observed effects have been constrained to specific conditions, such as when individuals were not previously exposed to the misinformation before its correction (Autry & Duarte, 2021), or among specific demographic groups, like individuals over 70 years old (Skurnik et al., 2005). Further evidence suggests that an explicit reminder of misinformation that repeats the initial information first and then explains how it is incorrect enhances the efficacy of corrections (Ecker et al., 2017; Wahlheim et al., 2020), especially when individuals are distracted during the correction process (e.g., listening to the radio

while driving) where repeated corrections are necessary to reduce misinformation reliance (Sanderson et al., 2022). This is likely due to the repetition of corrections facilitating the co-activation of both the misinformation and its correction, thereby making it cognitively easier to detect factual conflicts and integrate new information. As a result, stronger encoding and representation of the correction improves the subsequent recall of accurate claims (Sanderson et al., 2022).

Another useful strategy is to provide plausible alternatives in the correction. Studies have shown that the CIE can be reduced through the provision of an alternative account that explains why the misinformation was inaccurate (Johnson & Seifert, 1994; Kan et al., 2021; Nyhan & Reifler, 2015; Schwarz et al., 2016). For example, if a fire was thought to have been caused by negligently storing volatile materials, providing a causal alternative (i.e., "There is evidence for arson") is more effective than a simple negation (i.e., "There were no volatile materials found"). Because while a simple correction creates a gap in an individual's mental model, hindering their memory updating and knowledge integration process, a causal alternative explanation will fill in the gap and help complete their event recall picture (Ecker et al., 2022). Importantly, the alternative must be plausible, and ideally explains why the misinformation was offered and thought to be correct in the first place in order to successfully replace the misinformation (Lewandowsky et al., 2012). However, providing a plausible alternative might only reduce rather than eliminate the CIE, as belief in misinformation can increase over time despite the presence of an alternative (Rich & Zaragoza, 2020), probably due to individuals' lack of engagement in the effortful retrieval and monitoring required to assess the misinformation and its correction, which leads to incomplete revision of the mental model and the transience of individuals' revised beliefs.

To avoid the worldview backfire effect, worldview-congruent or self-affirming corrections should be provided. Research suggests that corrections should be tailored to their specific audience and framed in a way that confirms the audience's worldview (Lewandowsky et al., 2012). For example, reframing pro-environmental rhetoric in terms of purity, a moral value resonating primarily among conservatives, makes the information more acceptable among conservatives (Feinberg & Willer, 2013). Another way to make worldview-threatening corrections more palatable is through self-affirmation, which involves a message or task that bolsters one's feelings of self-worth (e.g., writing a brief essay about one's strengths and values). These practices are believed to help protect one's self-esteem and reduce the threat associated with the correction (Ecker et al., 2022). Indeed, research finds that self-affirmed individuals respond to worldview-challenging messages in a less defensive and more open-minded manner, leading to more favorable changes in beliefs and behaviors than individuals without self-affirmation (Carnahan et al., 2018; Sherman & Cohen, 2002).

Using graphical representations can also boost corrective message impacts by attracting attention, quantifying and disambiguating corrective information, and thereby facilitating comprehension and retention of corrections (Danielson et al., 2016; Dixon et al., 2015; van der Linden et al., 2014). Because people tend to prefer more simplistic messages than complex ones (Ecker et al., 2022; Walter et al., 2019), graphical representations accompanied with simple explanations can be more easily processed and integrated into memory. For example, Danielson et al. (2016) found that corrections with a combination of text, graph, and analogy were the most successful in changing misperceptions. However, the meta-analysis by Walter and colleagues (2019) discovered a weakened effect where graphical elements appear to attenuate the corrective effect such that visual corrections are less

effective than other formats of corrections across studies. Although it is not clear why the inclusion of visual elements may backfire, it is important to note that graphical representations should be easy to understand and are related to the critical misinformation to avoid misinterpretation or skepticism toward denials (Garrett et al., 2013).

Research controlling the retention interval of corrections has suggested that corrections should be provided immediately after exposure to misinformation (Ecker et al., 2015; Garrett & Weeks, 2013). The meta-analysis by Walter and Tukachinsky (2020) finds that the CIE is stronger when the correction is presented after a filler task compared to when the correction is presented without a delay. This is because once a mental model is constructed, the longer the misinformation is held, the more it is integrated into memory and the more difficult it is to eliminate it (Ecker et al., 2015). Therefore, compared to delayed corrections, real-time corrections are easier to integrate into the mental model and thus to update misperceptions. However, some exceptions exist where time of correction interacts with prior beliefs such that immediate corrections are less effective than delayed corrections when the corrections are contradicting the recipients' prior beliefs, because real-time corrections induce greater counterargument and can lead recipients to doubt the correction's credibility (Garrett & Weeks, 2013).

Last, corrections may be most effective when a combination of these strategies is used. Ecker et al. (2010) combined the provision of a specific warning (i.e., a warning message that explains the CIE in detail) and a plausible alternative in corrections, which revealed further reduction in the CIE than each one of the single strategies alone produced. Kan et al. (2021) also found that the combination of including an alternative and directly targeting the misinformation in corrective messages led to more successful elimination of the

CIE. While most research so far only examines one strategy or message design at a time, scholars have suggested conducting conjoint experiments where participants are exposed to multiple stimuli with many different combinations of interventions (Bak-Coleman et al., 2022; Motta, 2022).

In sum, to enhance corrective effects and reduce the CIE, corrections should be repeated to increase salience, presented with a plausible alternative explanation that explains why the misinformation was inaccurate, be consistent with the audience's worldview or be self-affirming of that worldview, be accompanied with graphical representations clearly and simply, be provided immediately after the misinformation when they are consistent with the recipients' prior attitude, and ideally should be reinforced with a combination of multiple strategies. While these recommendations may work to reduce the CIE and prevent some backfire effects, many other factors can influence the effectiveness of misinformation corrections. These come from different perspectives and will be discussed in Chapter 2.

Chapter 2 Research on Correction Features That Influence Their Effectiveness

Attempts to evaluate the effectiveness of misinformation corrections have revealed various features of corrections contributing to either attenuation or enhancement of the corrective effect. Generally, these influencing variables can be categorized into features pertaining to messages, sources, and audiences of corrections. But before delving into the different features, an operationalization of correction effectiveness is needed.

Operationalizations of Correction Effectiveness

Correction effectiveness is generally operationalized in three major ways in the literature—knowledge updating, belief or attitude change, and motivated behavioral intention. *Knowledge updating* refers to individuals' ability to recall facts about the tested event, and/or make inferential judgments based on the misinformation and its correction (Eva et al., 2021). A higher inferential reasoning score than the baseline (e.g., when there is no correction) reflects successful knowledge updating on the corrected information and thus correction effectiveness, which is a measurement commonly applied to CIE research (e.g., Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988).

For example, in the classic CIE laboratory paradigm, participants read a fictious news report that contains a critical piece of information (e.g., a recent warehouse fire was caused by volatile materials stored negligently in a closet) (Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988). This information is subsequently corrected (e.g., the closet was actually empty) or not corrected. Then, participants are asked to recall some basic facts about the event (e.g., "What time was the fire eventually put out?") and answer a series of causal inferential questions (e.g., "What could have caused the explosion?"). Recall accuracy scores are obtained from the number of correct facts recalled out of all of the fact questions as

potential attention check items, and inferential reasoning scores are obtained from the number of inferences based on misinformation out of all of the inferential questions. The inferential reasoning score is then used to measure the CIE.

It is also important to know how this operationalization of correction effectiveness unfolds over time. In some research exploring the duration of individuals' knowledge updating, participants are exposed to the misinformation and its correction, and tested either immediately after the exposure or after a delay of at least one week (e.g., Ecker et al., 2015; Ecker et al., 2020). Results show that corrective effects significantly wear off over time such that corrections are much less effective in reducing recall of false claims and reliance on misinformation in reasoning after a delay, because recollection of the correction can fade but familiarity of the misinformation remains relatively intact (Ecker et al., 2020). Research measuring post-correction knowledge updating over time also finds that updated knowledge is not durable because recall accuracy decreases and people continue to rely on misinformation in reasoning over one day following a correction (Rich & Zaragoza, 2020; Walter & Murphy, 2018). Furthermore, McIlhiney et al. (2022) found that the effectiveness of corrections remains temporally stable within an individual. In other words, an individual's susceptibility to the CIE after correction is relatively stable over time. When individuals are presented with a parallel CIE task after four weeks, there is a significant intra-class correlation between the first and the second CIE scores. These findings suggest that while knowledge of misinformation remains sticky in mind and continues to influence people's memory and judgement, corrected knowledge is not durable and thus fades over time, and such susceptibility to the CIE is relatively stable as an individual characteristic.

Outside of CIE research, *belief or attitude change* is another common way to operationalize correction effectiveness. Researchers often ask participants' belief in or agreement with a series of statements related to the misinformation after their exposure to the correction, compared to a control group where no correction is presented (e.g., Hameleers & Van der Meer, 2020; Vraga et al., 2019). For example, after viewing misinformation and a correction about the immigration issue, Carnahan and Bergan (2022) asked participants to rate the factual veracity of five claims (four false, one true) that are commonly circulated with regards to immigration in the U.S. (1 = definitely true to 4 = definitely false). Higher ratings of the false claims reflect greater belief accuracy or altered belief by the correction, as compared to the control group without correction. The belief (or in this case attitude change) can also be tested before and after exposure to correction with a retention interval, so that a difference score can be computed to capture change in beliefs within individuals (e.g., Lee, 2022; Swire, Ecker, et al., 2017).

Self-reported credibility perceptions of the misinformation can also indicate belief or attitude change. Credibility refers to the believability of a source or message on the part of the receiver (Hovland et al., 1953; O'Keefe, 1990). Therefore, perceived credibility of the misinformation can be considered as equivalent to belief in the misinformation. Typically, participants are asked to rate the perceived credibility of misinformation based on a Likert scale (e.g., 1 = Strongly disagree to 7 = Strongly agree) to indicate how much they agree that the information they viewed was accurate, believable, and authentic (e.g., Oeldorf-Hirsch et al., 2020; Vraga et al., 2020). Then, lower credibility perceptions of misinformation in the correction group compared to the no-correction control group will indicate effectiveness of the correction (e.g., Vraga et al., 2019). When an additional factor of the correction is

introduced (e.g., source credibility, individual differences) experimentally, lower perceived credibility of the misinformation in one condition than another indicates greater effectiveness of that specific version of the correction (e.g., corrections from a credible source are more effective than those from a less credible source) (Vraga & Bode, 2017a; Walter et al., 2020).

A within-subjects design that assesses credibility perceptions of the misinformation before and after exposure to the correction can also reflect correction effectiveness in altering misinformation belief (e.g., Tseng et al., 2022). It is worth noting that while a within-subjects design with pre- and post-correction measurements reduces error, it also risks priming participants about the information's veracity by asking their perceived credibility of the information before exposure to the correction, which may serve as a perceptual anchor for subsequent credibility ratings.

Some research also focuses on the favorability of recipients toward the misinformation or its source and investigates how corrections can decrease the likeability of either. In such studies, semantic differential items (Blankenship et al., 2015; Nan & Madden, 2012) are typically used to ask participants the extent to which they believe the misinformation is "harmful/beneficial," "foolish/wise," "bad/good," "unfavorable/favorable," or "useless/useful" on a Likert scale (e.g., Huang & Wang, 2022; Zhang et al., 2021). Research using statements by politicians examines how individuals' favorability toward the politician is changed by corrections of the politician's false statements (e.g., Agadjanian et al., 2019). Similar to the decayed accurate reasoning observed in longitudinal CIE research, studies on belief or attitude change also indicate that the positive effects of corrections tend to fade over time (e.g., Carey et al., 2022). However, corrections can have enduring effects when certain strategies are adopted, such as ensuring corrections are perceived as coming

from credible sources (Vraga & Bode, 2017a), are repeated and reinforced (Walter & Tukachinsky, 2020), or are detailed and align with individuals' existing knowledge or beliefs (Swire-Thompson et al., 2020).

Motivated behavioral intention to act on the tested issue can be viewed as an extended outcome of belief or attitude change, serving as a demonstration of correction effectiveness. Research has examined how corrections influence intentions to share misinformation (Chung & Kim, 2021; Yaqub et al., 2020), to engage in health behaviors such as vaccinations (Lee et al., 2022; Nyhan et al., 2014), to vote (Aird et al., 2018; Swire, Berinsky, et al., 2017), to reply to a correction (Martel et al., 2021), and to purchase products promoted by the correction (Tay et al., 2022). However, many studies show that corrections do not necessarily translate to behavior. While corrections may be effective in changing beliefs or attitudes, they seem to be less effective in motivating corresponding changes in behavioral intentions (e.g., Lee et al., 2022; Porter et al., 2022). This could be attributed to other influential factors, such as the emotional tone of the message, individuals' trust in online information, and their motivations for expressing opinions or interacting with others (Lee et al., 2022).

Overall, assessing knowledge updating as a metric for correction effectiveness delves into individuals' memory in terms of recall accuracy and inferential reasoning, shedding light on how reliance on misinformation persists from a cognitive perspective. Although studying the CIE of false information in fictional narratives offers valuable insights into general cognitive patterns, there remain gaps in the literature regarding the application of this knowledge updating approach to real-life misinformation and its correction. In practice, individuals' pre-existing beliefs or attitudes, coupled with various contextual and individual

factors, can significantly influence the effectiveness of corrections. Transitioning from the memory aspect, belief or attitude change serves as a more direct measure of correction effectiveness, soliciting individuals' self-reported beliefs, agreement, or attitudes toward misinformation. However, self-report bias exists and a consensus on a universally accepted and consistent measurement scale has yet to be reached. Meanwhile, the impact of corrections on motivated behavioral intention has been relatively modest or limited. However, it becomes more important in its practical application, as it focuses on influencing people's future behavior rather than mere cognitive or attitude adjustments.

By using real-life news issues and content, this dissertation will adopt the belief or attitude change method of operationalization to measure correction effectiveness. This will be done by assessing altered credibility perceptions of the misinformation in experimental groups compared to a no-correction control group, as well as between different experimental groups. In other words, if the experimental groups show reduced perceived credibility of the misinformation compared to the control group, it will indicate the general effectiveness of the corrections regardless of their varying features. Significant differences in perceived credibility of the misinformation among various experimental groups will indicate the greater effectiveness of corrections with specific features over others. A control group instead of a pre- and post-test design is used to avoid priming participants about the falsity of the misinformation by asking credibility perceptions of the misinformation before presenting its correction.

Features Influencing Correction Effectiveness

Various features of corrections influence their effectiveness. These features can be categorized into three main groups: message features, source features, and audience features.

These categories encompass a range of elements that can either enhance or diminish the corrective impact on misinformation. Understanding these features is crucial for developing more effective strategies to combat misinformation and improve the accuracy of public knowledge.

Message Features Influencing Correction Effectiveness

Message features of corrections can be further divided into *content variations* and *language variations* in the message. Content variations refer to adopting different correction strategies to vary the message content that leads to different levels of effectiveness. For example, research finds that *explanatory depth* impacts correction effectiveness such that a detailed explanation or factual elaboration in the correction (e.g., reinforcing new and credible facts that refute the misinformation) are more effective in countering attitudes and beliefs based on misinformation than is a simple negation (e.g., simply stating the misinformation is incorrect) (Chan et al., 2017; Jin et al., 2020; Swire, Ecker et al., 2017; Van der Meer & Jin, 2020). This finding is also consistent with the CIE paradigm where providing a new plausible alternative that explains why the misinformation was inaccurate reduces the CIE more than a simple refutation, because the detailed explanation fills in the gap of an individual's mental model (Ecker et al., 2022; Johnson & Seifert, 1994). It is further suggested that a detailed explanation should avoid repetition of the misinformation which could in turn strengthen misinformation persistence (Chan et al., 2017; Lewandowsky et al., 2012). However, exceptions exist whereas some research finds minimal effects of explanatory depth on correction effectiveness (Martel et al., 2021; Walter et al., 2020). Other influential factors, such as prior knowledge and belief certainty, are likely at play—detailed elaborations may be more effective in reducing belief in misinformation among individuals

with limited prior knowledge on the topic and low belief certainty (Walter et al., 2020). This underscores the importance of considering factors from multiple perspectives.

Other strategies that vary the content of corrective messages include using different framings for the message, such as *loss vs. gain*, *one-sided vs. two-sided*, *logic-focused vs. fact-focused*, and *narrative vs. non-narrative* framing. Research generally reveals that corrections with loss framing (e.g., Borah et al., 2021), two-sided framing (e.g., O'Keefe, 1999), logic-focused framing (e.g., Vraga et al., 2020), and narrative framing (e.g., Sangalang et al., 2019) are more effective in altering beliefs or attitudes based on misinformation than their counterparts, but with conditions. For instance, Borah et al. (2021) found that loss-framed corrections that highlighted negative consequences of not getting the HPV vaccine reduced misperceptions to a greater extent than gain-framed corrections that emphasized the benefits of the vaccine. Furthermore, the impact of *gain- vs. loss-framing* was contingent upon individuals' reflection levels, defined as their ability to connect new information to prior understanding. Individuals with low reflection exhibited reduced vaccine misperceptions when exposed to loss-framed corrections, whereas high-reflection individuals demonstrated low misperceptions regardless of whether they received gain- or loss-framed corrections. More moderators have been identified in other studies, including personal issue involvement (Gerend & Shepherd, 2007) and avoidance vs. approach motivation (Nan, 2012), adding constraints on the main effect of gain vs. loss-framing message design on correction effectiveness.

The advantages of *two-sided framing* were initially suggested by persuasion researchers who found that refutational two-sided messages elicited more positive attitudes and yielded greater credibility and attitude change than *one-sided* refutational messages in

persuasion contexts (Allen, 1991; O'Keefe, 1999). It is argued that refutational two-sided messages, which present both supporting and opposing arguments, demonstrate expertise and open-mindedness, leading to increased perceived source truthfulness and believability. Furthermore, by refuting the opposing opinions, they inhibit future counterarguments that might otherwise curb persuasiveness (Allen, 1991). When applied to the context of misinformation correction, Okuno et al. (2022) discovered that two-sided corrections that include both risks and benefits of vaccines were more effective in enhancing willingness to be vaccinated and regret of inaction than were one-sided corrections. However, the advantage of two-sided corrections diminishes with the introduction of additional factors. One-sided corrections are favored among individuals with limited prior experience on the issue (Wang & Huang, 2021) and higher initial misperceptions (Xiao & Su, 2021), as well as among those using on-line processing (i.e., forming immediate judgments upon encountering new information rather than rendering judgments later from memory as needed) (Carnahan & Garrett, 2020).

Similarly, *logic-focused* corrections, which highlight the rhetorical flaws of misinformation, have been found to be more effective in reducing misperceptions than *fact-focused* corrections, which provide accurate information refuting the misinformation (Schmid & Betsch, 2019; Vraga et al., 2020). For example, to debunk misinformation claiming that CO_2 is plant food and good for plants, a logic-focused correction illustrates the misleading technique used in misinformation that oversimplifies the logic between CO_2 and plants, while a fact-focused correction describes the scientific fact that plants need more than CO_2 to flourish, including the right temperature and amount of water (Vraga et al., 2020). As a result, logic-focused corrections effectively reduced misperceptions regardless of being placed

before or after exposure to misinformation, while fact-focused corrections only reduced misperceptions when they occurred after the misinformation. This might be because while the sharing of facts after exposure to misinformation is widely valued, logic-focused corrections can act as an inoculation pre-exempting individuals from potential misinformation attacks.

Narratives that tell stories are often more effective in altering beliefs or attitudes and behavioral intentions compared to non-narratives that simply present information (Braddock & Dillard, 2016; Murphy et al., 2015). This is largely because narratives provide a vivid and immersive experience of another person's perspective, making recipients less prone to ignore or avoid the message (Bilandzic & Busselle, 2013). Framing corrections as a narrative can also improve their acceptance and effectiveness in the context of misinformation (e.g., Sangalang et al., 2019), because a narrative can synthesize details about events and characters into a coherent cause-and-effect structure, which helps recipients update their mental model of the information in a coherent manner (Walter & Murphy, 2018; Zwaan et al., 1995). However, research also suggests that the effect of narrative corrections may be contingent on other factors. For example, Vafeiadis and Xiao (2021) found that highly issue involved individuals preferred non-narrative corrections with informational evidence because they had a direct stake in the issue. In contrast, low-involved individuals were more drawn to narrative corrections, as they were less interested in the issue and thus might find statistical information less engaging. People who use social media primarily for social interaction purposes are also more likely to accept and be influenced by narrative corrections, because they are more motivated to understand others' stories than those who use social media mainly for informational purposes (Lee, 2022).

Additional content variations of the corrective message include using new debunking tools such as "*truth scales*" or *multi-media formats*. A *truth scale* provides a clear visual indicator of the veracity of a claim such as a clear "True" or "False" fact-check label on social media. A finer scale can also judge the information as "mostly false" or "mostly true" to reflect more precision. Research finds that adding a truth scale to corrections increases their effectiveness in reducing belief in misinformation, especially among open-minded individuals, but only in non-political contexts (Amazeen et al., 2018). In political contexts, the inclusion of a truth scale can trigger reactance among partisans when it contradicts their prior beliefs or attitudes, weakening the corrective effect (Amazeen et al., 2018; Walter et al., 2019). *Using images or videos* as aids to corrective messages may also boost corrective impact by attracting attention, quantifying and disambiguating the corrective information, and as a result, facilitating comprehension and retention of corrections (Danielson et al., 2016; Dixon et al., 2015; van der Linden et al., 2014). Based on media richness theory, media that afford more cues and provide feedback are considered to have higher levels of media richness, which are preferred for messages containing complex information (Daft & Lengel, 1984). Therefore, corrections that include multiple media formats (e.g., text plus audiovisuals, etc.) may improve the evaluation and impact of the message. For example, Danielson et al. (2016) found that corrections with a combination of text, graph, and analogy were the most successful in revising misperceptions than a single textual correction. Tseng et al. (2022) also revealed the superiority of using video to correct textual misinformation by enhancing perceived credibility of corrections and motivating desirable actions.

Another set of message features vary the *language* of the message to influence correction effectiveness while keeping the content constant. For example, Bode et al. (2020)

explored whether the *tone* of corrections influence their effectiveness. By variating the tone of correction posts to be neutral by simply stating a fact refuting the misinformation (e.g., "This isn't true. Pasteurizing milk doesn't affect its nutrients"), or affirmational that shows empathy and affirms the original poster (e.g., "I know it can be super confusing, but this isn't true. Pasteurization doesn't affect the nutrients in milk at all"), or uncivil that insult the original poster (e.g., "Oh come on, don't be stupid. Everyone knows that pasteurizing milk doesn't affect its nutrients"), they found that all corrections reduced people's beliefs in misinformation regardless of the tone. It is possible that, unlike political topics, the health topic about pasteurizing milk is not salient enough to resist correction, and that an uncivil tone could even be problematic for more emotional and politicized issues (Bolsen & Druckman, 2018). Also, while the use of an uncivil tone by individual users on social media is more common and acceptable, such a tone could detrimentally impact the credibility of professional news outlets or fact-checking organizations who are expected to maintain a civil and professional tone, especially when correcting misinformation. Young et al. (2018) compared the effectiveness of humorous and non-humorous video corrections on a political issue (i.e., news delivered by a congressman on number of jobs created by a construction project) and found no advantages of humor over non-humor corrections. However, humor might be more effective in the context of highly polarizing issues, as it could potentially reduce counter-argumentation of attitude-discrepant corrections as suggested by Nabi et al. (2007). Some other findings on the effect of tone include a positive association between frequent civil corrections from friends on social media and a higher likelihood of successful corrections in the context of a political election (Heiss et al., 2023), and a negative effect of using a tough demanding tone in COVID-19 related corrections (e.g., "xxx is wrong!") on

their effectiveness (Zhang et al., 2022), but more research is needed to test the effect of tone in other contexts.

A related message feature is *language intensity* that is defined as "the quality of language which indicates the degree to which the speaker's attitude toward a concept deviates from neutrality" (Bowers, 1963, p. 345). It describes how intense, strong, assertive, extreme, emotional, and vivid people perceive the language of a message to be (Hamilton & Stewart, 1993). Messages with high language intensity typically contain intensifiers such as "great" instead of "good," "very large" instead of "large," and "nice!!!" instead of "nice" to increase the strength or emotional appeal of the language (Burgers & de Graaf, 2013). Some research finds that messages using intense language can enhance attitude change by increasing message strength (i.e., making the information more vivid) and by being perceived as clearer and more logical (Craig & Blankenship, 2011; Hamilton & Hunter, 1998; Hamilton & Stewart, 1993; McEwen & Greenberg, 1970). However, other studies suggest that language intensity may negatively impact message evaluation and persuasiveness, potentially reducing perceived credibility of corrections and their effectiveness by activating reactance among recipients (Kim et al., 2017; Quick & Considine, 2008; Xue, 2021). These conflicting results warrant further investigation of language intensity, which is a major focus of this dissertation, along with an examination of other influential factors on correction effectiveness, as explained in Chapter 3.

Emotional tone is another message feature that significantly influences message persuasiveness. To differentiate it from emotional state, which is an audience feature that describes how the individual is feeling, emotional tone of a message refers to what emotions or the extent of emotions (i.e., emotiveness) the message is trying to appeal to or express

from the use of emotion-evoking language. Research indicates that negative emotional appeals are effective in changing attitudes, intentions, and behaviors when the audience perceives high self-efficacy in avoiding the harm (Tannenbaum et al., 2015). Moreover, the emotional valence of news headlines can overshadow the impact of source cues on judgments of news content (Baum & Abdel Rahman, 2021), possibly because emotional language attracts more attention, and is more salient and accessible from memory (Lang et al., 1995). By incorporating emotional appeals in a correction, Sangalang et al. (2019) found that narratives with emotional corrective endings were more effective in reducing misbeliefs than a simple correction without emotions. A correction utilizing multiple emotional appeals, combining fear and anger, was also more successful than a single, discrete emotion in reducing various misinformation outcomes including attitudes, beliefs, and intentions. However, while the inclusion of emotions may bolster persuasiveness under certain conditions, their prevalent use in misinformation may also imply deceitfulness, potentially leading recipients to disengage from emotionally-driven corrections (e.g., Xue, 2021).

Additionally, *information readability* (i.e., the level of effort required to understand the information) can influence the acceptance of a correction, because providing information that is easy to read is necessary to facilitate recipients' positive attitudes toward the correction (Sui & Zhang, 2021). By varying the complexity of wording in a correction, Wang et al. (2022) revealed a positive association between information readability and correction acceptance, especially among people with high cognitive abilities. The authors also found a positive effect of *argument quality* (i.e., the strength or plausibility of persuasive argumentation) on correction acceptance, which is more of a content variation feature that varies the quality of evidence and reasoning that a correction uses to refute misinformation.

It is important to note that the literature reviewed in this dissertation only focuses on features of corrections, but not of misinformation, although features of misinformation can also interact with correction features to influence a correction's effectiveness (e.g., Ecker et al., 2011; Hameleers, 2022; Tseng et al., 2022). Overall, the research literature on various message features influencing correction effectiveness has yielded mixed results, with significant effects observed under specific circumstances. This highlights the importance of investigating the combined effects of multiple features (e.g., message and audience features) from different perspectives.

Source Features Influencing Correction Effectiveness

In addition to message features, source features of corrections are equivalently important as influential factors or even determinants of correction effectiveness under some circumstances. One of the most examined source cues is *source credibility*, which refers to the believability of a source of some information. In terms of persuasion, the efficacy of a communication is commonly presumed to rely significantly on the source delivering it. Not surprisingly, research shows that message persuasiveness increases with perceived source credibility—high-credibility sources are more persuasive than low-credibility sources in changing audience beliefs, attitudes, or behaviors (Hovland & Weiss, 1951; Pornpitakpan, 2004). Given that a correction is a form of persuasive message intended to change an audience's beliefs (Robertson et al., 2020), its effectiveness may also hinge on how credible the source is perceived to be. While people tend to judge misinformation to be more accurate when it comes from sources perceived as more credible (Traberg & van der Linden, 2022), corrections from high-credibility sources are also evaluated positively and have more efficacy in reducing reliance on and beliefs in misinformation (Vraga & Bode, 2017a; Walter

et al., 2020; Walter & Tukachinsky, 2020). For instance, by varying the source of a correction to originate from either another social media user or an expert source (e.g., the CDC), researchers found that a single correction from the CDC was sufficient to reduce misperceptions about the origins of the Zika virus, whereas a single user correction was not as successful in altering misbeliefs (Vraga & Bode, 2017a). A meta-analysis on the impact of corrections on health misinformation on social media also revealed significantly stronger effects of corrections from expert sources than from non-experts (Walter et al., 2020).

The significant role of perceived source credibility is further emphasized in a paper that investigates the effectiveness of corrections from various sources (i.e., professional fact-check organizations, mainstream news outlets, social media platforms, artificial intelligence, and crowdsourcing), such that credibility perceptions of a correction's source are positively associated with its effectiveness in reducing belief in misinformation (Liu et al., 2023). Moreover, by assessing perceived source credibility across multiple dimensions (i.e., expertise, trustworthiness, objectivity, benevolence, and transparency) inspired by the MAIN model (Sundar, 2008), another paper examining the same set of sources of corrections unveiled distinctions in certain dimensions of source credibility among different sources (Metzger et al., 2023), which may explain the varied effectiveness of these sources in more nuance. Specifically, professional fact-check organizations were perceived as the most expert and credible source likely guided by the authority heuristic (i.e., the perception that the source is an expert on the topic at hand), and thus more effective in correcting health-related misinformation. AI was rated relatively high on objectivity and trustworthiness, likely guided by the machine heuristic (i.e., an assumption that machines are more objective than humans), facilitating its efficacy in reducing misbeliefs (Metzger et al., 2023). Some dimensions are

also found to be more influential than others. For example, receiving a correction from a source high in trustworthiness, rather than expertise, reduced individuals' reliance on misinformation when making references (Ecker & Antonio, 2021; Guillory & Geraci, 2013; Pluviano et al., 2020). This suggests that trustworthiness might play a more important role in impacting correction effectiveness than expertise does, probably because some people lack general trust in experts (e.g., political, academic) and in science nowadays.

However, the effectiveness of social corrections (e.g., corrections from friends or peers on social media) are more complex to evaluate. On one hand, individuals may consider misinformation and corrections from their peers as more credible than other correction sources due to the influence of social ties. People tend to trust information from those who resemble them, even if these individuals lack specific expertise (Wang et al., 2008). Indeed, a meta-analysis reveals that correcting misinformation conveyed by peers is more difficult than correcting misinformation disseminated by news agencies (Walter et al., 2020). Indeed, collective social corrections of political misinformation motivated by group norms and self-presentation purposes have proved to be successful in facilitating information verification processes among social groups (Kligler-Vilenchik, 2022). On the other hand, social corrections may be perceived as less credible and thus less effective in debunking health misinformation compared to authoritative sources (e.g., health agencies or physicians), especially when it comes to disaster information or emerging crises (Schultz et al., 2011; Wogalter, 2006). Research shows that corrections from the CDC and a news media agency (i.e., Reuters) are more successful in altering people's misperceptions of health crises than corrections from their peers on social media (Van der Meer & Jin, 2020). Vraga and Bode (2017b) further revealed the necessity of providing an outside source (such as the CDC or a

professional fact-check organization) in social corrections to effectively reduce misbeliefs about the outbreak of a virus, whereas social corrections lacking an outside source failed to mitigate misperceptions compared to the no-correction control group.

Considering the substantial impact of both message and source features on correction effectiveness, it is worthwhile to inquire about their relative importance when both are present in a correction. Research shows that elevating people's focus on either the source or the message can amplify the influence of source or message features on persuasiveness. For instance, Albarracín et al. (2017) instructed participants to evaluate either the message arguments or source credibility of a political advertisement, and then recall either the arguments or the source attributes before assessing their attitudes toward the advertised political party. They found that when individuals focused on the message arguments, persuasion increased when strong arguments were delivered by a noncredible communicator. Conversely, when attention was directed toward source credibility, persuasion was amplified when credible sources presented weak arguments. This finding is further corroborated in the context of misinformation, where emphasizing the information source reduced accuracy ratings of plausible news headlines from uncredible sources, and increased accuracy ratings of implausible headlines from credible sources (Dias et al., 2020). This work offers valuable insights for misinformation corrections, emphasizing the need to highlight either or both their credible sources and high-quality messages to maximize their effectiveness.

The Elaboration Likelihood Model (ELM) can also be employed to address this question. Based on the ELM, individuals differ in how carefully and extensively they think about a message and the position it is advocating (i.e., varying levels of elaboration), which is influenced by their motivation and ability (Cacioppo & Petty, 1984; Wagner & Petty,

2011). When motivation and ability are high (i.e., high elaboration level), people are more likely to think via the central route where they thoughtfully assess the merits of the information by focusing on the message content and quality. When motivation and ability are low (i.e., low elaboration level), people are more inclined to go down the peripheral route where they rely on cues and heuristics (e.g., source cues) to process the information without examining the message thoroughly. Therefore, moderated by individual characteristics, people may process message cues and source cues of corrections through different routes. For example, Wang et al. (2022) used computational methods to investigate how the central route and the peripheral route cues in the ELM influence the acceptance of misinformation corrections and the moderating effect of cognitive ability. They found that individuals with low cognitive abilities relied more on source credibility as a peripheral cue and argument quality as a central cue to accept rebuttals, while individuals with high cognitive abilities relied more on information readability as another central cue to accept rebuttals. This highlights the value of examining message, source, and audience features to understand correction effectiveness.

Audience Features Influencing Correction Effectiveness

Various individual-level audience characteristics also help to elucidate why corrections prove effective for certain individuals but not for others. Given the cognitive effort necessary to comprehend and process corrections of misinformation, *cognitive abilities* of the audience have emerged as significant factors influencing the acceptance of corrections and adjustments in misbeliefs. Specifically, cognitive abilities such as *working memory (WM) capacity* and *verbal intelligence* have been recognized as predictors of susceptibility to misinformation effects. Limited WM capacity heightens the risk of experiencing the CIE due

to incomplete updating, manipulation, and removal of information from WM (Brydges et al., 2018). Corrections are also more successful in altering attitudes toward targeted political information among individuals with higher *verbal intelligence* than lower verbal intelligence, even when controlling for open-mindedness and right-wing authoritarianism (De keersmaecker & Roets, 2017).

Analytic thinking, which refers to a disposition to engage in effortful, deliberative thinking and inhibit intuitive, heuristic-driven information processing (Frederick, 2005), is another cognitive ability that has been studied. Studies indicate that individuals with higher levels of analytic thinking exhibit significantly more favorable attitudes toward corrections of misinformation compared to those with lower levels of analytic thinking (Allen et al., 2021; Epstein et al., 2022; Lyons et al., 2020). Martel et al. (2021) also found that analytic thinking (measured by the Cognitive Reflection Test) and actively open-minded thinking predicted increased acceptance of corrective information, regardless of correction style. This is likely because analytic thinking fosters skepticism toward conspiratorial concepts (Pennycook et al., 2015), and individuals with enhanced analytic thinking abilities are better equipped to discern and navigate through the complexities of information, making them more resilient to the influence of misinformation and more amenable to correction.

Apart from cognitive abilities, the *pre-existing worldview* of the audience can significantly influence the acceptance of corrections. Research shows that pro-attitudinal corrections aligned with people's *prior belief* are often more effective in altering misbeliefs than counter-attitudinal corrections unaligned with one's prior belief (Banas et al., 2022; Walter et al., 2019; Walter & Tukachinsky, 2020). This can be explained by motivated reasoning and the confirmation bias (Kunda, 1990; Taber & Lodge, 2006) that are associated

with the worldview backfire effect. Motivated reasoning is also believed to be related to analytic thinking, whereby analytic thinking facilitates the formation and maintenance of beliefs congruent with one's identity, leading to an increased likelihood of motivated reasoning and opinion polarization (Thorson & Li, 2021). Consequently, individuals with elevated levels of analytic thinking may exhibit a greater tendency toward motivated reasoning when confronted with counter-attitudinal corrections, mitigating the potential benefits of analytic thinking in overall correction efficacy. This underscores the necessity of taking individual's prior beliefs into account, particularly in contexts involving correcting political or contentious issues.

However, here again some studies found conflicting results where corrections are more effective in reducing misbeliefs (Carnahan & Bergan, 2022) and misinformation engagement intentions (Lee et al., 2022) among individuals who harbored prior beliefs unaligned with the corrections. Although some of this could be argued as a floor effect, other variables should be considered to explain the mixed results. For example, Vraga and Bode (2017b) found that while a correction from another user failed to reduce misperceptions, corrections from an expert source (i.e., CDC) were effective in correcting misbeliefs, especially among people with higher initial misperceptions, which was also replicated in their later study (Vraga et al., 2022). This suggests that source credibility acts as a moderating factor, influencing individuals' decisions to accept incongruent corrections. It is likely that low-credibility sources serve as a discounting cue, enabling information processors to dismiss their messages (Hovland et al., 1949). Conversely, messages from high-credibility sources are less easily discounted and can facilitate successful attitude change,

even among those influenced by motivated reasoning (Albarracín & Vargas, 2010; Jamieson & Hardy, 2014).

While people hold prior beliefs toward an issue, their *belief certainty* varies, which can affect their willingness to accept a correction and adjust their attitudes. Measuring belief certainty together with belief accuracy is important in addressing some validity issues in the operationalization of misperceptions. When individuals provide inaccurate responses to factual questions in surveys, it is either due to firmly held incorrect beliefs, or merely bad guessing on unfamiliar items they are uncertain about (Pasek et al., 2015). This uncertainty stems from insecurity about one's knowledge. Individuals are often aware of their own level of knowledge and respond with low certainty when providing incorrect answers (Graham, 2018). Based on the Differential Informedness Model by Li and Wagner (2020), the presence, the certainty, and the accuracy of one's prior belief will lead to different levels of informedness among individuals, which influence the success of corrections in terms of belief updating. Upon testing the model, they found that misinformed individuals (i.e., people who held incorrect prior beliefs with high certainty) were less likely to change their misbeliefs than were uninformed individuals (i.e., people who held no prior beliefs). The ambiguous individuals (i.e., people who held prior beliefs with low certainty despite correct or incorrect), however, were not different from the uninformed individuals in their extent of belief updating after exposure to corrections on the immigration issue. Carnahan and Bergan (2022) also investigated how belief certainty impacts the effectiveness of corrections, but discovered significant corrective effects even among those who held inaccurate prior beliefs with high certainty (on immigration issues). More research is needed to further explore the

influence of belief certainty across different issues and contexts (e.g., politicized versus nonpoliticized misinformation).

Another audience feature regarding one's existing worldview is *personal issue involvement*. Individuals are involved with a topic when it is of intrinsic importance, has personal meaning, and/or has significant consequences for them (Petty & Cacioppo, 1986). Individuals with high involvement tend to allocate more cognitive resources to evaluate the quality of messages or arguments, whereas individuals with low involvement may rely on heuristic cues (e.g., source credibility) to process information (Petty et al., 1981; Petty et al., 1983). High involvement with an issue may facilitate one's tendency toward motivated reasoning when encountering attitude-inconsistent information. Therefore, while low-involved people may accept a correction more easily (especially when it is from a high-credibility source) when it contradicts their prior beliefs, high-involved people may express stronger attitudes and a willingness to only adopt attitude-congruent corrections and to discount contradicting ones (Petty & Cacioppo, 1986; Vafeiadis & Xiao, 2021). Research on the CIE also finds that misinformation with higher relevance to individuals has larger continued influence on their memory and reasoning (Jin et al., 2022). While more empirical evidence is lacking in the context of misinformation, personal issue involvement is an important variable that requires more attention and investigation in future research.

The effectiveness of misinformation corrections may also vary among groups with diverse *demographics* such as age, gender, education, political ideology or partisanship, and political sophistication. Research finds that some groups are more vulnerable to misinformation than other groups in certain contexts including older populations (e.g., Guess et al., 2019), males (e.g., Epstein et al., 2022) or females (e.g., Chen et al., 2015), less

educated populations (e.g., Gurgun et al., 2024), Republicans (e.g., Holman & Lay, 2019), and more politically sophisticated individuals (e.g., Vitriol et al., 2022).

Age can impact the efficacy of corrections given its indication of diminished cognitive abilities, lack of familiarity with the Internet, and limited knowledge of how to evaluate online information among individuals aged 65 or older (Guess et al., 2019). Older adults are also more susceptible to the CIE and more consistently hold inaccurate beliefs even after exposure to corrections, particularly over long intervals (Swire, Ecker et al., 2017). However, some studies reveal that older people tend to challenge misinformation more (Gurgun et al., 2024) and share misinformation less than young people (Epstein et al., 2022), although the mean age of participants in those studies was below 50.

When it comes to *gender*, one study discovered that women are less inclined to share misinformation compared to men (Epstein et al., 2022), perhaps because prior research suggests that women are more cautious in online information seeking (Lim & Kwon, 2010). Nevertheless, another study found that more female college students have shared and intend to share misinformation on social media platforms than males have given their higher rate of social media sharing in general (Chen et al., 2015). Men are also more likely to challenge misinformation than women (Gurgun et al., 2024), probably because men are more willing to express their opinions online (Bode & Vraga, 2021). Considering the mixed results, no decisive conclusions can be drawn from the current literature on gender differences in influencing the effectiveness of misinformation corrections.

Individuals with higher levels of *education* are more likely to accept and engage with corrections regarding health misinformation on social media (Bode & Vraga, 2021; Gurgun et al., 2024; Walter & Murphy, 2018), because education is in general positively associated

with opinion expression intentions (Gurgun et al., 2024) and more training in critical thinking and media literacy (Bode & Vraga, 2021). However, more educated political partisans may exhibit greater motivated reasoning tendency, and thus more resistance to accept attitude-incongruent corrections of political misinformation (Nyhan et al., 2013). Another study on misinformation about COVID-19 suggests that although a knowledge gap exists where more educated individuals display higher levels of knowledge about COVID, belief in misinformation did not significantly differentiate across education levels (Gerosa et al., 2021). Thus, education can serve as a double-edged sword in combating misinformation, with no clear indication that higher levels of education are inherently beneficial unless they are specifically focused on fostering discernment between truth and fake news or between correction techniques.

Political ideology or partisanship has been identified as an influential audience feature in extensive misinformation research. Given the current highly polarized political context, people's prior beliefs or attitudes on an issue are often heavily affected by their political partisanship, leaving those with a strong political affiliation particularly vulnerable to misinformation (Carey et al., 2022). A partisan difference exists whereas Republicans or conservatives appear to be more susceptible to attitude-consistent misinformation and more likely to engage in motivated reasoning when confronted with ideology-challenging corrections, compared to their Democratic or liberal counterparts (Walter et al., 2019). For instance, in their study on correcting misinformation about election fraud among individuals from both parties, Holman and Lay (2019) found that Republicans were more likely to correct their false beliefs when the correction source was aligned with their ideology rather than contradictory or neutral (e.g., professional fact-check organizations). In contrast,

Democrats were more flexible in changing their misbeliefs by correction sources that are ideologically inconsistent or neutral. This difference is possibly caused by Republicans' increasing negative perceptions of professional fact-checking organizations (Shin & Thorson, 2017), their greater proneness to closed-mindedness, and less intention to seek out counter-attitudinal information (Barberá et al., 2015).

Political sophistication, which includes aspects of political engagement, knowledge, and interest, is found to facilitate political motivated reasoning (Miller et al., 2016; Taber & Lodge, 2006). Compared to political novices, those with higher levels of political sophistication are more capable and willing to defend their political loyalties, and to respond more negatively to counter-attitudinal information that attacks their favored political allies (Taber & Lodge, 2006; Vitriol et al., 2022). Consequently, political sophistication may exaggerate individuals' belief in ideology-congruent misinformation and disbelief in out-party misinformation (e.g., Jenke, 2023), as well as their counter-arguing tendency toward fact-checks that challenge their beliefs (e.g., Young et al., 2018). However, in the case of character-based misinformation (i.e., misbelief about a politician's personal integrity), rather than policy-based misinformation (i.e., misbelief about an issue such as climate change), political sophisticates (as measured by political knowledge) are better able to discern real news from fake news, regardless of their partisan consistency with the news (Vegetti & Mancosu, 2020). This underscores the importance of examining the role of political sophistication across various contexts of misinformation.

Additional audience features that have been investigated in the literature include *tolerance for negativity*, *media use, media literacy*, and *emotional states* of recipients (e.g., Ecker et al., 2022; Fridkin et al., 2015; Jones-Jang et al., 2021; Li et al., 2022; Weeks, 2015;

Xiao et al., 2021; Yu et al., 2023). Research finds that individuals with less *tolerance for negativity* are more likely to accept fact-checks correcting the false claims made in negative advertisements that attack political candidates, despite their partisanship (Fridkin et al., 2015). Increased social media use for news consumption is also associated with lower levels of knowledge and more fake news beliefs (Gerosa et al., 2021; Pennycook & Rand, 2019; Tandoc et al., 2018). Greater news *media literacy* (i.e., higher media locus of control and need for cognition) can lead to more fake news literacy (i.e., the ability to detect fake news), which in turn encourages people to take corrective action when encountering misinformation (Huber et al., 2022). And *emotional states of recipients* can act as a mediator such that the presence of a correction that argues a health crisis is more severe than the misinformation stated makes people feel less hopeful, and more confused and fearful, which then leads to attitude change more aligned with the correction (i.e., increased perceived crisis severity) (Van der Meer & Jin, 2020). However, reliance on emotion (e.g., Martel et al., 2020) and arousal of anger (e.g., Weeks, 2015) or happiness (e.g., Forgas & East, 2008) can also make people more vulnerable to misinformation in the first place. Understanding the intricate interplay between emotions and information processing is a key area for comprehending how individuals respond to corrections and misinformation in diverse contexts.

Gaps in Addressing Features from All Perspectives

Although there is extensive literature on many features from each perspective (i.e., message, source, and audience), researchers generally look at these factors in isolation. Limited research examining more than one factor simultaneously mainly focuses on the effectiveness of corrections with different message styles, or from different sources, among individuals with congruent or incongruent prior attitudes (Boukes & Hameleers, 2022;

Garrett & Weeks, 2013; Holman & Lay, 2019; Vraga et al., 2022; Vraga et al., 2019) or other personality traits (Fridkin et al., 2015; Martel et al., 2021; Porter et al., 2022). However, research that considers all three (message, source, and audience factors) in the same study is almost entirely lacking from the literature.

Yet to understand how and when fact-checks are effective, it is important to look at the combined effect of these factors because real-world fact-checks involve multiple cues simultaneously. For example, an individual encountering a fact-check Tweet is exposed to both the message itself in terms of the Tweet's content, format and language style, and the source (i.e., the person or organization who posts the Tweet). The acceptance or rejection of the fact-check is also influenced by the individual's states and traits (e.g., emotion, need for cognition). Therefore, it is only when taking message, source, and audience factors into account collectively that we can gain a more comprehensive and realistic picture of fact-checking effectiveness.

As a notable example, one study attempted to investigate relationships between factors from all three perspectives. Employing computational communication methods, Wang et al. (2022) analyzed a large number of posts and comments on social media related to fake news about COVID-19, and draws upon the elaboration likelihood model (ELM) to reveal distinct effects within each information processing route. Specifically, they found positive effects of central route elements (i.e., information readability and argument quality) on acceptance of misinformation corrections, representing message factors. There were also effects of peripheral route elements (i.e., source credibility, including authority and influence) on correction acceptance, which highlights source factors. Interestingly, source authority negatively affected acceptance, whereas source influence, indicated by number of followers

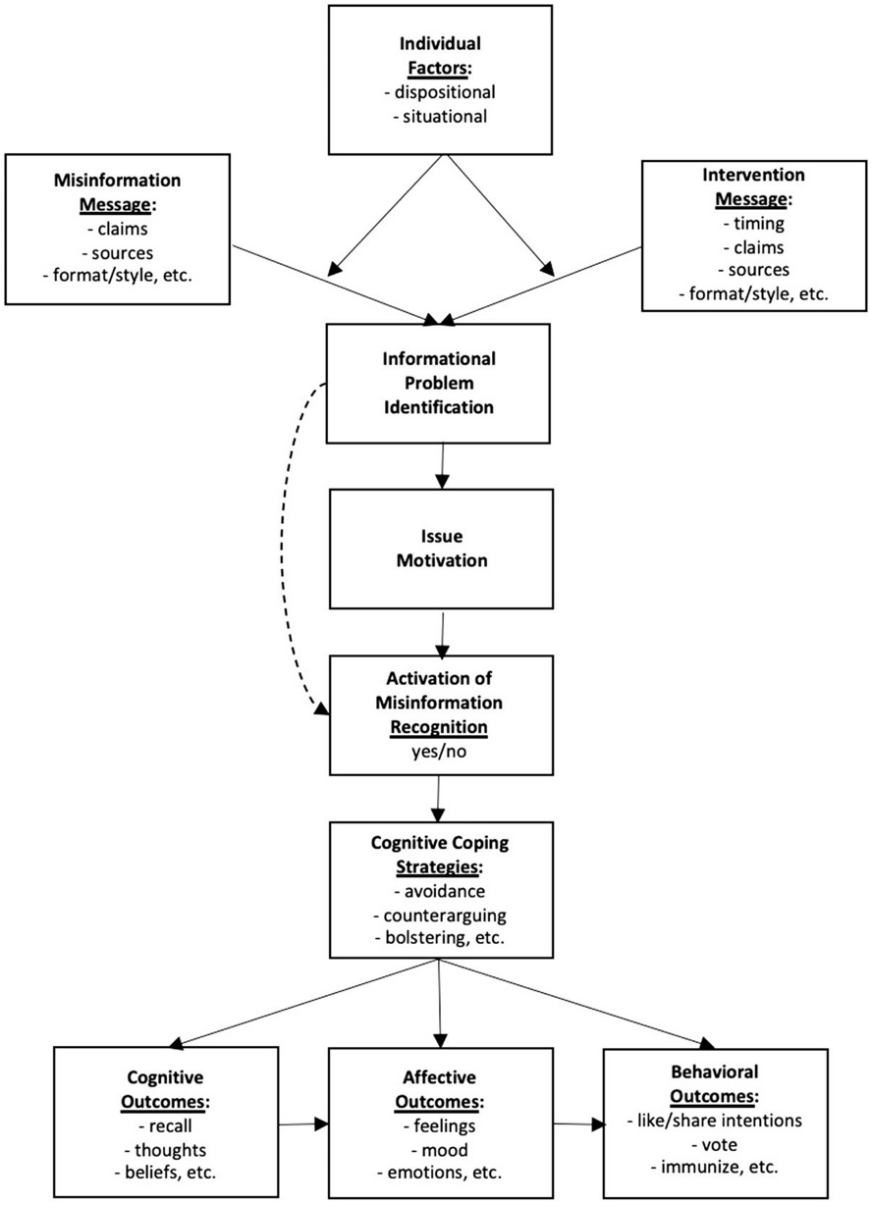
of the source, yielded a positive impact. In terms of audience factors, moderation effects of individuals' cognitive abilities on these relationships were observed. While those with lower cognitive abilities relied more on source credibility and argument quality to accept corrections, individuals with higher cognitive abilities focused on evaluating information readability in making acceptance decisions. Although the study did not investigate interaction effects between the factors, this endeavor demonstrates the potential to assess the effectiveness of misinformation corrections in a more holistic manner.

Echoing the need for a more comprehensive approach to misinformation research, Amazeen (2024) introduced the Misinformation Recognition and Response Model (MRRM) and encouraged researchers to explore how various perspectives intertwine and contribute to a broader understanding of misinformation. The model offers a framework for examining both the antecedents to and consequences of misinformation recognition. It theorizes that individual factors interact with the content of misinformation and corrective messages to influence cognitive processes in dealing with (mis)information. Individual factors include dispositional factors that are relatively stable over time (e.g., demographics, need for cognition, media literacy), and situational factors that vary upon circumstances (e.g., prior attitude, involvement, and knowledge about the topic). Message factors include the claims, sources (which arguably represent source factors), and format or delivery styles of the misinformation and corrections. The model further proposes that by taking into account one's own situational and dispositional characteristics and evaluating the misinformation or corrections based on their message features, individuals can determine the extent to which their beliefs align with the message, leading to either the activation of misinformation recognition when they agree with the correction or no activation when they agree with the

misinformation. Then, cognitive coping strategies (e.g., avoidance, counterarguing) are employed that result in different outcomes including cognitive, affective, and behavioral outcomes (Amazeen, 2024). See Figure 1 for a visual depiction of the MRRM. While individuals might not consistently adhere to the suggested cognitive framework or consider their individual and message factors when encountering misinformation or corrections, it is inspiring for researchers to thoroughly examine these factors in combination and to embrace comprehensive perspectives in misinformation studies.

Therefore, the central aim of this study is to address the gap in misinformation research identified above by investigating factors from all three perspectives by taking message, source, and audience factors into account to evaluate the effectiveness of misinformation corrections. Specifically, this study selects language intensity of the correction as a central message feature to assess how corrective messages with different levels of language intensity vary in their effectiveness in altering belief in misinformation (measured by perceived credibility of the misinformation). This relationship is further explored under the influence of source credibility as a source feature, alongside issue involvement and attitude discrepancy as audience features to provide a more comprehensive understanding of the effectiveness of corrections in this specific context. In the next chapters, empirical research and theoretical foundations of each factor will be discussed and hypotheses based on these foundations will be developed (Chapter 3). Chapter 4 will delineate the methodology used to test the hypotheses. The results will be presented in Chapter 5 and then discussed in Chapter 6.

Figure 1. Conceptual Process of the MRRM by Amazeen (2024)



Chapter 3 Study on the Effect of Language Intensity on Correction Effectiveness

The persuasion effect of a message depends on the content of the message, the source of the message, and individual-level characteristics of the audience (Hovland et al., 1953). This dissertation will examine aspects of each of these three factors, with an emphasis on language intensity as a message factor, for reasons that are articulated in this chapter.

Language intensity, a feature of the message content that refers to the degree of specificity and emotionality conveyed through language (Hamilton et al., 1990; Hamilton & Stewart, 1993), can affect how a message is perceived by the audience. Subtle linguistic variations in the message (e.g., "This is fantastic!" vs. "This is nice") can arouse different levels of attention or emotions, influencing the audience's emotional response and the persuasiveness of the communication. Variations in language intensity can also connote partisan journalism practices that suggest reporting bias or even inaccurate facts reported (Burgers & de Graaf, 2013). As such, misinformation corrections with different levels of language intensity may imply varied untruthfulness of the corrected information and convey the strength of the correction source's attitude (e.g., "This news is misleading" vs. "This news is fake!"). This has the potential to lead to varying levels of attitude change and, consequently, impact the effectiveness of misinformation corrections.

While language intensity has been extensively studied in various persuasion contexts (e.g., health promotional messages, political campaigns), it is underexplored in the context of misinformation. News reports are generally expected to maintain a neutral and professional tone to avoid suggesting bias or opinionatedness. However, fake news often employs extreme language and evokes strong emotions, which can lead to widespread dissemination and complicate correction efforts. In the social media environment, emotional messages,

particularly those inciting hatred or division, can easily go viral (Brady et al., 2017; Vosoughi et al., 2018). In such a landscape, the principle of maintaining journalistic neutrality may no longer be effective in combating misinformation. It is possible that appropriately using language intensity in corrections could "fight fire with fire," thereby increasing public exposure to and acceptance of corrections.

In this dissertation, *source credibility* is selected as the major feature of the correction's source. Credibility is a critical factor in persuasive communication because it affects the audience's trust and acceptance of the information presented. When a source is perceived as credible, the audience is more likely to accept the message (Hovland et al., 1953). Applied to fact-checking, source credibility may thus enhance the correction's effectiveness in changing attitudes and beliefs. What is less clear from the literature is the effect of language intensity on source credibility. For reasons that will be discussed later in this chapter, source credibility is likely to play a significant role in the relationship between language intensity and correction effectiveness.

Additionally, two audience factors of interest to this research are *attitude discrepancy* (i.e., the degree to which a correction's verdict disconfirms an individual's prior attitude toward the veracity of the misinformation being corrected), and *issue involvement*. Prior attitude is crucial in this research because it significantly impacts the effectiveness of corrections in addressing misinformation (Walter et al., 2019). For instance, when a correction's verdict starkly contrasts with an individual's preexisting beliefs, it may influence how they perceive the correction due to cognitive dissonance (Festinger, 1957). When individuals encounter information that challenges their preexisting beliefs, they may experience discomfort or tension, leading them to dismiss or undermine the correction to

maintain their original viewpoint. This reaction could be exacerbated by the intensity of the language used in the correction. Intense language might provoke a stronger emotional response, making individuals more resistant to accepting the correction if it conflicts with their prior attitudes. Thus, the interaction between attitude discrepancy and language intensity may significantly impact the effectiveness of misinformation corrections.

Issue involvement, defined as the degree of importance, personal meaning, and consequences an issue has to a person (Petty & Cacioppo, 1986), is also known to interact with language intensity to affect persuasion (Kronrod et al., 2012). The varied effectiveness of corrections across different topics (e.g., politics, health, etc.) further suggests the necessity to take issue involvement into account (Vraga et al., 2019; Walter & Murphy, 2018). Moreover, a combined effect of issue involvement and attitude discrepancy may occur such that increased issue involvement enhances the effectiveness of pro-attitudinal corrections with intense language but reduces the effectiveness of counter-attitudinal corrections with intense language (Petty & Cacioppo, 1986; Vafeiadis & Xiao, 2021) for reasons to be discussed later in this chapter.

Drawing these threads together, this study aims to investigate the impact of source, message, and audience factors by proposing a model of correction effectiveness that focuses on the effect of language intensity, termed the "LICE" Model (Language Intensity and Correction Effectiveness Model), which is informed by the information processing theory of language intensity effects (Hunter et al., 1984) and language expectancy theory (Burgoon et al., 1975). The LICE model examines perceived language intensity as a central message feature, perceived source credibility as a source feature, and attitude discrepancy and issue involvement as audience features.

The message feature of language intensity was prioritized as the focal variable to study for two reasons. First it is easier to implement into corrections than it is for fact-checkers to influence either source or audience features, and thus message features have the most potential both in terms of practical application and real-world effects in fighting misinformation. While controlling or altering audience members' credibility perceptions of the source of correction messages or audience characteristics—such as issue involvement or prior attitudes—is likely impossible (or would require immense time and effort for fact-checkers to discern) in the wild, modifying a correction's message features is both feasible and cost-effective for fact-checkers to implement. It is also interesting and important to study language intensity as a central feature because it can theoretically affect correction effectiveness in a positive *or* a negative direction. While high-intensity language may improve correction effectiveness by making the correction more vivid and engaging, it may also reduce effectiveness by decreasing perceived credibility of the correction. Examining this variable in the new context of misinformation correction can thus help clarify mixed findings from past research and reveal contextual differences in the impact of language intensity. Thus, while taking source and audience features into consideration, this dissertation centers investigating language intensity of corrections as the critical feature of correction effectiveness, highlighting its theoretical and practical significance in the socially-important context of combatting misinformation.

Moreover, the LICE model will be tested in the context of social media, where misinformation stories and corrections are presented as social media posts, specifically on Instagram in this study. According to Pew Research (2024), 53% of U.S. adults at least sometimes get news from social media platforms, following news websites or apps and

search as other primary news sources. Misinformation is also particularly prevalent on social media, where it can spread rapidly due to the platforms' social functions (Buchanan, 2020). In fact, large percent of users see untrue content almost every time they use various platforms such as Facebook, TikTok, X, and Instagram, and more than half of social media users admit to sometimes sharing news without verifying the facts (Vigderman, 2024). In response, multiple efforts have been made to combat misinformation on social media. Mainstream news outlets post fact-checks for their own content or that of others. Professional fact-checking organizations also conduct and share fact-checks, while social media platforms monitor and flag content that might be misleading or false. Given the widespread influence and challenges of misinformation on social media, examining the effectiveness of corrections in this context is both timely and highly relevant.

Importantly, the model includes two variations: one for sources *with* prior credibility perceptions and one for sources *without* prior credibility perceptions (see Figures 2 & 3 on page 54 for the model variations). This distinction is made because the presence or absence of prior source credibility perceptions likely triggers different cognitive processing of the correction message, and thus affects correction effectiveness. Specifically, the LICE model proposes that while corrections using intense language may reduce perceived message and source credibility and ultimately, reduced correction effectiveness, when they are from low-credibility or no-prior-credibility sources, their effectiveness may increase under certain conditions. These conditions include when the corrections come from familiar, high-credibility sources, align closely with the recipient's prior attitudes, and address topics with high issue involvement. The following sections explicate each model component.

Language Intensity, Perceived Message Credibility, and Correction Effectiveness

Language intensity was initially defined as "the quality of language which indicates the degree to which the speaker's attitude toward a concept deviates from neutrality" (Bowers, 1963, p. 345). However, this definition is problematic because it requires defining neutrality, which is difficult to achieve due to varying perspectives and subjective interpretations of both speakers and audiences (Liebrecht et al., 2019). Hamilton and Stewart (1993) suggest that language intensity can be expressed through two language features: specificity and emotionality. Specificity refers to "the degree to which a source makes precise reference to attitude objects in a message," and emotionality is "the degree of affect expressed in the source's language" (p. 231). High-intensity language is characterized by emotional and specific words that make the information contained in the message more concrete and vivid. Various intensifiers can be used to increase language intensity. For example, words can be replaced with a more extreme version (e.g., *great* instead of *good*), words can be added to exaggerate (e.g., *very large* instead of *large*), and typographical elements can be used to emphasize (e.g., *nice!!!* instead of *nice*) (Renkema, 1997). Typically, messages with high language intensity contain imperatives and controlling terms (e.g., must, should, ought) while low-intensity messages include terms that emphasize autonomous actions (e.g., could, might want to, worth) (Miller et al., 2007).

While various intensifiers can be used to manipulate the intensity of language, audiences may perceive messages to have different levels of language intensity. To address this, Hamilton and Stewart (1993) developed the perceived language intensity scale as a manipulation check to see if the intensified version of a message is indeed perceived as more intense than a version without intensifiers. This study will use perceived language intensity

both as a manipulation check and as a continuous predictor variable for correction effectiveness (rather than high or low language intensity as a categorical variable). This approach allows for a more nuanced understanding of language intensity from the perspective of individual perceptions.

Early research suggests that messages using more intense language provide a greater motivation for recipients to adopt the message-advocating position because they are perceived as clearer and more logical (Hamilton & Hunter, 1998; McEwen & Greenberg, 1970). High-intensity language also heightens emotional arousal and is more likely to capture audience members' attention, leading recipients to process the message more deeply, which in turn can enhance its persuasive effect (Bowers, 1963). By making the message more vivid and engaging, intense language helps the message remain in conscious thought for a longer period of time, especially when message discrepancy is low (i.e., when individuals either agree with or only slightly disagree with the message position) (Hamilton & Stewart, 1993).

However, more recent persuasion research finds opposite results where message evaluation and persuasiveness are negatively associated with the use of high-intensity language. For example, Dai et al. (2022) found that low-intensity or non-assertive messages facilitated the intention of green consumption behavior of individuals, compared to high-intensity or assertive messages. Winter et al. (2015) also found that recipients tended to be skeptical about statements that were presented as overly certain, thus reducing its effectiveness in persuasion compared to less assertive language. This inverse relationship may be explained by *reactance* theory, which says that when individuals' perceived behavioral freedoms are threatened or reduced, people are aroused psychologically and motivated to restore the threatened freedoms by reacting negatively to the stimuli and

refusing to comply (Brehm, 1966). For example, intense messages advocating regular exercise with overtly persuasive language (e.g., language containing many imperatives and controlling terms) tend to be perceived as a greater threat to freedom than non-intense messages, resulting in higher levels of anger and more negative assessments of message fairness (Miller et al., 2007). Many other studies have shown that intense or forceful language can amplify reactance, leading to reduced message evaluation and persuasive effects in public health or environmental campaigns (Kim et al., 2017; Quick & Considine, 2008; Quick & Stephenson, 2007).

Nevertheless, persuasive messages in theses contexts are likely to differ from corrections in the misinformation context. On one hand, persuasive messages in public health campaigns often aim to promote proactive, long-term behavioral changes (e.g., exercising regularly), which requires carefully framing the message to avoid triggering resistance. In these campaigns, the goal is to encourage voluntary compliance and self-motivated behavior change, making non-assertive, low-intensity language more effective. In contrast, fact-checks or corrections in the misinformation context are designed to directly confront false beliefs. Unlike public health campaigns, where the focus is on motivating behavior, fact-checks are more focused on providing immediate clarification or debunking falsehoods. This may require presenting information in a more direct or authoritative manner, to reestablish facts and counteract the potential harm of misinformation. As a result, fact-checkers may be motivated to show some level of assertiveness to ensure the false claims are clearly and effectively addressed (Graves et al., 2015). For example, professional fact-checking organizations often use assertive labels to emphasize the falsehood of public remarks (Ferracioli et al., 2022). In this case, intense language may project confidence of the source

and make the correction more effective. However, the use of high-intensity language in corrections can also backfire, especially when the intensity of the message is perceived as a threat to personal autonomy or belief.

To assess the effect of language intensity on misinformation correction effectiveness, credibility perceptions of the correction message are likely to serve as key mechanisms in explaining the underlying cognitive processes. As an important criterion for both message and source evaluation, credibility is a multidimensional concept consisting of dimensions such as expertise, trustworthiness, completeness, bias, objectivity, goodwill, and transparency (e.g., Hovland et al., 1949; Rieh & Danielson, 2007). *Message credibility* (Appelman & Sundar, 2016), defined as "an individual's judgment of the veracity of the content of communication" (p. 63), can be influenced by source as well as by non-source factors, such as the structure of the messages and the medium of delivery. Although persuasion research generally reveals negative effects of high language intensity on both perceived message credibility and persuasiveness (Miller et al., 2007; Kim et al., 2017), more research is needed on the effect of language intensity in the context of misinformation. It is likely that, similar to other persuasive messages, the perceived credibility and effectiveness of corrective messages will decrease with the use of more intense language. Only one study, by Xue (2021), has examined this variable in the misinformation context and found that corrections using high-intensity labels (e.g., "fake") resulted in lower perceived message credibility and acceptance of the correction compared to those using low- or mid-level intensity labels (e.g., "inaccurate" or "wrong"). This effect may occur because people generally expect corrective messages to be professional and objective, and high-intensity language might imply a lack of

objectivity, thereby reducing the perceived credibility and persuasive impact of the message (Xue, 2021). It is also important to note that Xue's study examined unfamiliar sources only.

Theoretical foundations that help explain the negative effect of language intensity on message credibility perceptions and persuasiveness include Burgoon's language expectancy theory (LET). This theory posits that people develop normative expectations about appropriate language use in persuasive messages, and deviations from these expectations can influence their attitudes and behaviors toward the message (Burgoon et al., 1975). High-intensity language is often associated with deceptive messages, which frequently use assertive, expressive, and directive speech to try to enhance their persuasive power (Iswara & Bisena, 2020; Zhou et al., 2003). Thus, when a corrective message—typically expected to be professional and objective—contains high-intensity language, it may violate these expectations, leading to reduced credibility perceptions and, consequently, diminished effectiveness of the correction.

Moreover, given the proliferation of fake news and increasing polarization of partisan media agencies, some members of the public may hold a skeptical view of mainstream or other media sources. This skepticism can be especially triggered by intense or hyperbolic language, further reducing the acceptance and perceived credibility of both news and corrections, especially when they are from untrusted sources. As a result, people may be less likely to change their beliefs about misinformation when presented with corrections featuring high language intensity. Therefore, when taking no other factors (e.g., prior source credibility, message discrepancy, issue involvement) into account, the LICE model proposes a preliminary generic relationship between language intensity and correction effectiveness based on the LET. This relationship is worth exploring given the application of this theory to

a new context in order to see if the theory is supported in the case of misinformation correction. Thus the first hypothesis is:

H1: Increased perceived language intensity reduces correction effectiveness (i.e., increases perceived credibility of the misinformation) by decreasing perceived message credibility of the correction (path a_1b_1 in Figure 2 or path ab in Figure 3).

Figure 2. Proposed Model for the Effects of Perceived Language Intensity, Message Credibility, and Source Credibility on Perceived Credibility of Misinformation for *Unfamiliar* Sources

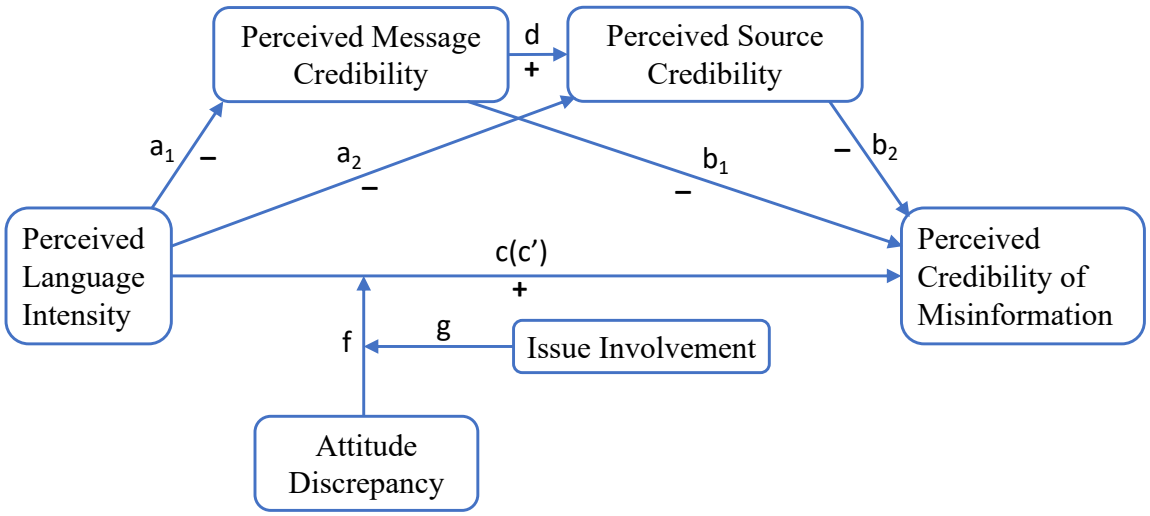
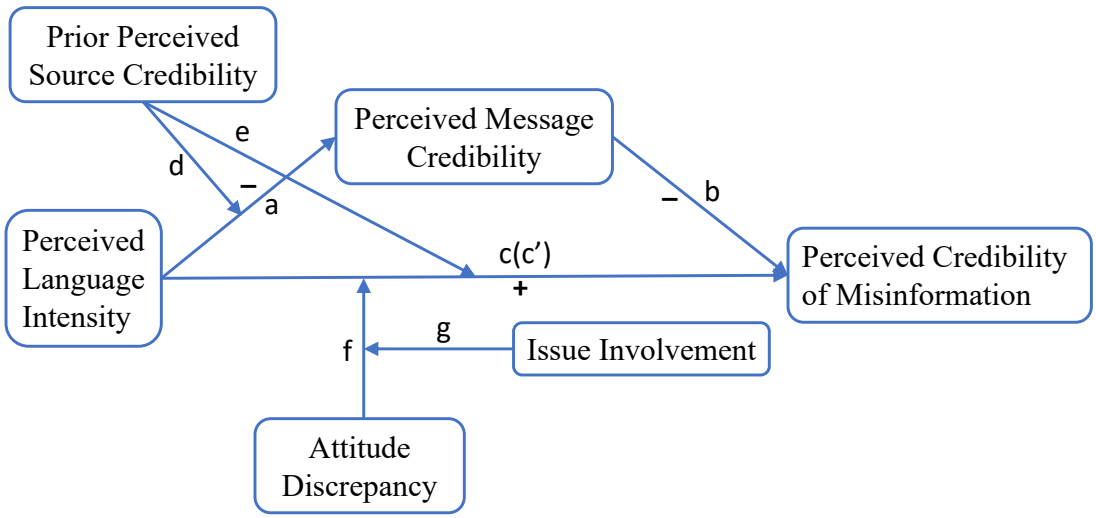


Figure 3. Proposed Model for the Effects of Perceived Language Message Credibility, and Prior Source Credibility on Perceived Credibility of Misinformation for *Familiar* Sources



Source credibility refers to the believability of a source of information. It is important because people tend to rely on source credibility cues to process and evaluate information (Sundar, 2008), as this approach requires fewer cognitive resources and effort than deliberately evaluating the information content (Metzger et al., 2010). However, depending on individuals' familiarity with or prior knowledge of the source, the effect of language intensity on source credibility and the overall effectiveness of corrections may vary. On one hand, when source cues are accessible and prominent—such as when people are familiar with and hold strong prior attitudes toward the source—the source effect may override the impact of message features in determining the correction's effectiveness (Chaiken & Maheswaran, 1994). For example, even if *FOX News* employs intense language in its corrections, partisans who inherently trust *FOX News* as a source might still perceive their correction messages as credible, thereby making them effective in reducing misbeliefs.

On the other hand, when people are less familiar with a source—such as novel sources like AI and crowdsourced fact-checking agents—or when the source cue is ambiguous or confusing (e.g., messages attributed to multiple sources through reposting), the absence of clear source cues may lead individuals to focus more on the message content, where linguistic features become more salient and influential in their evaluation of the message (Appelman & Sundar, 2016). In situations where there is no established credibility perception of the correction source because the source is unfamiliar, linguistic features such as language intensity may play a more significant role in shaping impressions of the source's credibility, which in turn may affect the correction's effectiveness.

Given increasing polarization between partisans and the rise of new correction sources that are less familiar to the U.S. public, such as AI and crowdsourcing corrections,

source credibility has become particularly important in studying the effectiveness of corrections. The LICE model, therefore, adapts to different contexts by incorporating perceived source credibility in the relationship between perceived language intensity and correction effectiveness—serving as a mediator when the source is unfamiliar (see Figure 2) and as a moderator when the source is familiar (see Figure 3), as explained below.

Perceived Source Credibility as a Mediator for Unfamiliar Sources

In the case of an unfamiliar correction source, evaluation of the message may provide evidence for the source's credibility, such that lower message credibility indicates lower source credibility, and vice versa (Slater & Rouner, 1996). The idea that source and message credibility mutually influence each other has been termed "credibility transfer" by Schweiger (2000). While most research focuses on how source credibility affects message credibility (e.g., Swire, Berinsky et al., 2017; Traberg & van der Linden, 2022), it is likely that people form impressions of unfamiliar sources based on their perceived message credibility derived from judging the message's content and qualities as suggested by Tormala et al. (2006). For example, in the context of political communication, voters often encounter statements from lesser-known candidates. If a candidate delivers a message that is factually sound, logically consistent, and aligns with the voters' values, the message's credibility can enhance the candidate's perceived credibility, even if the voters had no prior knowledge of the candidate. Research has shown that when the message content is strong and aligns with a receiver's pre-existing beliefs, it can positively impact the perceived credibility of the unknown source (Pornpitakpan, 2004).

As a result, a decline in source credibility resulting from decreased message credibility is likely to reduce the effectiveness of a correction. Research indicates that

message persuasiveness increases with perceived source credibility—high-credibility sources are more persuasive than low-credibility sources in changing audience beliefs, attitudes, or behaviors (Hovland & Weiss, 1951; Pornpitakpan, 2004). In the context of misinformation correction, expert sources with high credibility perceptions tend to be more effective than non-expert sources in reducing beliefs in and sharing of health misinformation (Vraga & Bode, 2017a; Walter et al., 2020; Zhang et al., 2021). This suggests that perceived credibility of a source plays an important role in whether a correction is accepted or rejected. Therefore, it is likely that:

> **H2:** For unfamiliar sources with no existing credibility perceptions, increased perceived language intensity reduces perceived message credibility of the correction, which in turn lowers perceived source credibility of the correction, ultimately leading to reduced correction effectiveness (i.e., increased perceived credibility of the misinformation) (path a_1db_2 in Figure 2).

Increased perceived language intensity may also influence perceived source credibility directly. Research shows that the use of intense or controlling language lowers assessments of sociability and trustworthiness of a source (Miller et al., 2007), and increases perceived extremity of the source's position (Hamilton & Stewart, 1993). It also results in greater freedom threat perceptions and thus reactance, which leads to source derogation (Ma & Miller, 2022). Communicators who use high-intensity language in their messages have been found to be rated as less credible than those who use less intense language (Bradac et al., 1980; Hamilton & Hunter, 1998). A meta-analysis by Hamilton and Hunter (1998) showed that although language intensity increases perceived dynamism of a speaker (i.e., how forceful, active, and intense the speaker seems to be), it negatively affects ratings of

source expertise and trustworthiness. LET may also help explain this because people expect a fact-checking agent to be neutral and objective, and usage of high intensity language may indicate extremity and partisanship of the source's position, which violates recipients' expectations and may lead to lower source evaluations in terms of credibility.

While research on the effect of language intensity on perceived source credibility in the context of misinformation correction is scant, other investigations of political issues have revealed consistent results. For example, political candidates using low-intensity language received higher scores on source credibility than did candidates using high-intensity language (Clementson et al., 2016). Therefore, it is likely that high-intensity language used in a correction decreases the perceived credibility of the source, ultimately reducing the correction's effectiveness. Another mediating path is thus hypothesized as:

> **H3**: For unfamiliar sources with no existing credibility perceptions, increased perceived language intensity reduces correction effectiveness (i.e., increases perceived credibility of the misinformation) by decreasing perceived source credibility of the correction (path a_2b_2 in Figure 2).

(Prior) Perceived Source Credibility as a Moderator for Familiar Sources

Nevertheless, in the case of a familiar source, once a correction source that individuals have prior credibility perceptions about is introduced, linguistic features such as language intensity are less likely to alter pre-existing attitudes toward the source, usurping the indirect mediating paths described above for unfamiliar sources (i.e., H2 and H3). This is because pre-existing attitudes about the source are difficult to change due to people's motivated reasoning where counter-attitudinal information is likely to be disregarded (Kunda, 1990). Therefore, recipients exposed to corrections from a familiar source they perceive as

credible may discount the negative effect of high-intensity language because they trust the source. In other words, the positive influence of high source credibility may override the negative effects of high language intensity, thereby not undermining the effectiveness of corrections. This suggests that pre-existing credibility perceptions of the correction source may act as a moderator, independently influencing the relationship between perceived language intensity and correction effectiveness.

The information processing theory of language intensity effects (Hunter et al., 1984) posits that attitude change is a three-way interactive function of message intensity, discrepancy (i.e., the inconsistency between the recipient's prior attitude and the message), and source credibility. When discrepancy is constant, attitude change is enhanced with high intensity and high credibility; but is inhibited with low intensity and low credibility. Based on this theory, Hamilton et al. (1990) found that language intensity enhanced the persuasiveness of a high-credibility source but inhibited it for a low-credibility source, while having no effect for a moderate-credibility source. Likewise, Buller et al. (2000) found that sun safety messages from credible sources (i.e., schools and pediatricians) were more persuasive to parents when they contained more rather than less intense language. Moreover, Burgoon et al. (1975) found the combination of a high-credibility source with high-intensity language, as well as a low-credibility source with low-intensity language, led to greater attitude change than other combinations.

Another theory that explains the varying persuasiveness of intense messages from high- vs. low-credibility sources is language expectancy theory. Specifically, LET suggests that people develop expectations about message strategies others will use in persuasive attempts (e.g., fear appeals, opinionatedness, language intensity) based on socio-cultural

norms and preferences (Burgoon et al., 2002; Burgoon et al., 1975; Burgoon & Miller, 2018). Positive expectancy violations occur when (a) the persuasive source exceeds expectations by using message features that surpass the normative bandwidth positively, and (b) when a source initially evaluated negatively unexpectedly conforms to the receivers' values and norms. In both cases, the receivers are likely to change their attitudes and behaviors in the desired direction of the source. Conversely, negative expectancy violations occur when a persuasive source uses message strategies that fall outside the bandwidth of socially acceptable behavior in a negative direction, leading to no attitude change or changes opposite to the advocated direction.

High-credibility sources are expected to have a wider bandwidth of acceptable persuasive strategies, allowing them more freedom and flexibility in enacting a larger variety of message strategies without violating expectations negatively. On the other hand, low-credibility sources, due to their narrower bandwidth of acceptable communication, risk being evaluated negatively and decreasing compliance when employing more aggressive strategies (e.g., fear appeal, highly intense language) (Buller et al., 2000). As a result, high-credibility sources that use intense language and low-credibility sources that use less intense language may achieve greater persuasiveness than other combinations through positive expectancy violations. In the misinformation context, corrections from a high-credibility source (e.g., a reputable organization like Factcheck.org) may be given more leeway to express confidence and assertiveness in their language (e.g., "This is a lie!") without risking their effectiveness, leading to positive attitude change. Conversely, corrections from a low-credibility source may be perceived as committing a negative expectancy violation when using highly intense

language.[1] Thus, for sources where prior credibility attitudes exist, the LICE model proposes that:

> **H4:** For familiar sources with existing credibility perceptions, increased perceived language intensity leads to greater correction effectiveness (i.e., reduced perceived credibility of the misinformation) when prior perceived source credibility is high, but reduces effectiveness when source credibility is low (path e in Figure 3).

Similarly, prior perceived source credibility may also moderate the effect of perceived language intensity on perceived message credibility such that:

> **H5:** For familiar sources with existing credibility perceptions, increased perceived language intensity leads to greater perceived correction message credibility when prior perceived source credibility is high, but reduces correction message credibility when source credibility is low (path d in Figure 3).

Attitude Discrepancy as an Audience Feature

The discrepancy between a correction's verdict and recipients' beliefs in the veracity of the message being corrected may influence how perceived language intensity affects correction effectiveness. In political contexts, pro-attitudinal corrections (i.e., debunking misinformation that supports an opposing ideology) often have a stronger impact on reducing misbeliefs compared to counter-attitudinal corrections (i.e., debunking misinformation that supports the recipient's own ideology) (Walter et al., 2019). Similarly, Boukes and Hameleers (2022) found that corrections were particularly effective in lowering agreement with and

[1] Rather than using unfamiliar sources as the low credibility manipulation, as done in some prior studies, this research assumes that unfamiliar sources have *no* prior credibility perceptions, whereas familiar sources can be perceived as varying from low to high in their credibility perceptions.

perceived credibility of misinformation among individuals whose prior attitudes were congruent with the correction's verdict. This can be explained by motivated reasoning and confirmation bias, where people seek, interpret, and believe information that aligns with their existing attitudes, while discounting counter-attitudinal information, even in the face of concrete evidence and mounting facts (Kunda, 1990; Taber & Lodge, 2006).

However, these results might also be due to a floor effect—individuals whose prior beliefs consistent with the correction may already hold accurate beliefs and are thus less likely to believe the misinformation in the first place, making it seem as though the correction is more effective than it actually is. When examining changes in belief in misinformation, research shows that people exposed to counter-attitudinal corrections experience more significant reductions in misbeliefs than those exposed to pro-attitudinal corrections (Liu et al., 2023). Therefore, instead of comparing groups *with no attitude discrepancy* at all (i.e., individuals who disbelieved the misinformation and held beliefs congruent with the correction) to those *with attitude discrepancy* (i.e., individuals who believed the misinformation and held beliefs incongruent with the correction), it may be more meaningful to compare groups with *low attitude discrepancy* (i.e., individuals whose beliefs slightly oppose the correction) to those with *high attitude discrepancy* (i.e., individuals whose beliefs moderately or completely oppose the correction). Under this operationalization, attitude discrepancy should only contain positive values where higher values indicate greater discrepancy between the correction's verdict and recipients' beliefs, with zero indicating no discrepancy. Including a small negative value may also be meaningful, because individuals whose beliefs are only slightly congruent with the correction in the beginning might feel validated by it, resulting in significantly lower perceived

credibility of the misinformation than the control group. This operationalization of attitude discrepancy is novel in the misinformation literature and offers a more meaningful way to test correction effectiveness among populations most susceptible to misinformation, while avoiding the floor effect for those who are less likely to believe misinformation in the first place.

According to Hunter et al.'s (1984) information processing theory of language intensity effects, a receiver reacts to a communication or messages by comparing each argument or assertion about the object being described with their own beliefs. Arguments that agree with people's own beliefs will not change their attitude toward the object, but assertions different from their beliefs may induce corresponding attitude change if the receiver yields to the argument. The basic information processing model for attitude change assumes a linear relationship between discrepancy and change: the greater the discrepancy between prior attitude and the message, the greater the attitude change (Anderson, 1959). However, alternative models predict nonlinear relationships, suggesting that beyond an inflection point in the discrepancy curve, further discrepancy may lead to counterarguments that prevent attitude change (Eagly, 1974; Hamilton & Stewart, 1993).

When taking language intensity and credibility into account, the information processing theory of language intensity effects (Hamilton & Stewart, 1993; Hunter et al., 1984) posits that with source credibility held constant, attitude change and source derogation increase with both language intensity and attitude discrepancy, because greater intensity makes the source's position seem more extreme or discrepant from the recipient's prior attitudes, leading to greater attitude change, but only if discrepancy functions monotonically. Nevertheless, nonlinear models suggest that once discrepancy reaches a certain point, high

language intensity may trigger resistance and inhibit attitude change. In the context of misinformation corrections, high discrepancy toward a political issue may be especially common due to partisan polarization and stereotypes toward ideological outgroups, compared to non-politicized issues (e.g., environmental or health topics) where individuals may hold neutral attitudes and low discrepancy with corrections. As a result, a correction with high discrepancy from the recipient's prior beliefs about the veracity of the misinformation, combined with high perceived language intensity, may have little to no success in correcting misbeliefs.

To extend the information processing theory of language intensity to the context of misinformation, the LICE model proposes that attitude discrepancy moderates the relationship between perceived language intensity and correction effectiveness. Specifically, the model asserts that when source credibility is held constant, perceived language intensity and attitude discrepancy interact to affect correction effectiveness such that: (1) when attitude discrepancy is high, intense language may heighten resistance to the message, thereby reducing correction effectiveness; and (2) when attitude discrepancy is low, intense language might not trigger resistance, but rather improve the persuasiveness of the message by making the message look more vivid and engaging, leading to greater correction effectiveness.

> **H6**: When source credibility is held constant, increased perceived language intensity reduces correction effectiveness (i.e., increases perceived credibility of the misinformation) when attitude discrepancy is high, but improves correction effectiveness when attitude discrepancy is low (path f in Figure 2 & 3).

Issue Involvement as an Audience Feature

Individuals are involved with a topic when it holds intrinsic importance, personal meaning, and significant consequences (Petty & Cacioppo, 1986). Research indicates that the more involved people are with an issue, the more likely they are to comply with messages promoting that issue (A. Clark, 1993; T. Clark, 1998), particularly in contexts such as pro-environmental (Cleveland et al., 2005) and health-promoting messages (Marshall et al., 2008), although this is likely true only when they agree with the message. Based on the LET, issue involvement may shape linguistic expectations because intense language conveys urgency and demands attention, aligning with the expectations of a highly involved issue. Consequently, a combination of high language intensity and high issue involvement should positively violate expectations, leading to greater attitude or belief change in the desired direction of a low-discrepancy message. Conversely, for issues with low involvement, intense language might be perceived as unexpectedly and overly aggressive or inappropriate, potentially leading to lower compliance due to its perceived excessive forcefulness. In such cases, low-intensity or less assertive language is more effective for gently persuading the audience. This is supported by Kronrod et al. (2012), who found that in various environmental contexts, recipients were more willing to comply with a persuasive message with assertive phrasing on issues they were involved with, but preferred non-assertive appeals for issues on which they were less involved (Baek et al., 2015; Wieluch, 2017).

However, the effect of issue involvement on message persuasiveness is likely more complex in the context of misinformation, because people tend to hold stronger prior attitudes toward politicized issues than toward neutral topics, such as recycling or exercising, which are commonly studied in persuasion research. Individuals are also more likely to

counterargue with a message that contradicts their prior attitudes, especially on issues they are highly involved with. Therefore, it would be too presumptuous to predict that individuals will always prefer high-intensity language corrections for issues they are more involved with and low-intensity language corrections for issues they are less involved with, without considering their prior attitudes toward the misinformation's veracity.

Research shows that while people with low issue involvement may readily accept counter-attitudinal information, those with high issue involvement express stronger attitudes and willingness to only adopt attitude-congruent information while discounting contradictory information (Petty & Cacioppo, 1986; Vafeiadis & Xiao, 2021). Thus, corrections with high attitude discrepancy may be more effective in changing misbeliefs among low-involved recipients than among those with high involvement. Bringing language intensity back into the picture, it is also reasonable to assume that when language intensity is high, corrections with high attitude discrepancy will be more effective among individuals less involved with the issue than those who are highly involved, because the latter may perform a greater degree of scrutiny on the message, engage in motivated reasoning, and perceive the intense language as another discounting cue (aside from high discrepancy) to reject the correction. By contrast, when language intensity is high and attitude discrepancy is low, highly-involved individuals may be more receptive to the correction than less involved individuals, because they are less motivated to counterargue, and more likely to be persuaded by the intense language, leading them to correct their misbeliefs on an issue they deeply care about for the sake of self-education. These predictions are also consistent with nonlinear information processing models, where attitude change increases with language intensity and low attitude discrepancy up to a certain point, but decreases with increased intensity and high discrepancy beyond that

point. Therefore, it is hypothesized that there will be a three-way interaction between perceived language intensity, attitude discrepancy, and issue involvement in predicting correction effectiveness:

> **H7**: The effect suggested in H6 will be strengthened as issue involvement increases such that when source credibility is held constant, increased perceived language intensity reduces correction effectiveness when attitude discrepancy is high, but improves correction effectiveness when attitude discrepancy is low, particularly among individuals who are highly involved with the issue (path g in Figure 2 & 3).

Combined Effects of Message, Source, and Individual Features

While the LICE model primarily investigates how language intensity as a message feature influences the effectiveness of corrections, it also examines this effect in the context of controlling for either individual characteristics (H1 through H5) or source credibility (H6 and H7). However, it is important to consider whether these elements—message, source, and individual features—might interact to influence correction effectiveness in a combined manner. This raises the question: How do these factors work together to shape the impact of misinformation corrections? To address this, the following research question is posed:

> **RQ1:** What are the combined effects of perceived language intensity of the correction, perceived source credibility of the correction, attitude discrepancy, and perceived issue involvement on correction effectiveness?

This research question aims to explore the complex interplay between the message's linguistic style, the credibility of the source delivering the correction, and individual characteristics such as the audience's existing attitudes and level of involvement with the issue at hand. Understanding how these factors interact could reveal a more nuanced picture

of what makes a correction more or less effective, highlighting potential interactions and cumulative effects that extend beyond the influence of each factor in isolation.

Chapter 4 Method

This research proposes two variations of the LICE model to assess the effect of perceived language intensity on correction effectiveness for sources with or without prior credibility perceptions, while taking perceived message credibility, attitude discrepancy, and issue involvement into account (see Figures 2 & 3). To test the model, an online experiment was conducted.

Participants

1630 participants were recruited from Prolific with monetary compensation. According to recent research comparing the data quality collected through five commonly used platforms (i.e., MTurk, Prolific, CloudResearch, Qualtrics, and SONA), Prolific and CloudResearch provide data of higher quality than other platforms and with the least cost paid per high-quality respondent (Douglas et al., 2023). After excluding people who completed the study in under 200 seconds (n = 15), who confessed to having googled relevant information during the study (n = 11), and who failed the attention check (i.e., "Select 'Somewhat agree' as the answer to this question") (n = 23) and a manipulation check item that asks them to identify who the fact-check they saw today was from (n = 79), a total sample of 1509 participants were included in the data analyses. Power analysis reveals that with a medium effect size of .26 (see Table 2), and a significance level at .05, the sample size of 1509 is more than sufficient to power the study to an optimal level (close to 1).

Experimental Design and Procedure

This study used a 2 x 2 x 2 x 2 between-subjects design. Factors are issue involvement (high vs. low), language intensity (high vs. low), attitude discrepancy (high vs. low), and source (unfamiliar source vs. familiar source), including a control condition where

participants saw no correction.[2] A pilot test employing 60 participants from Prolific was conducted before the main experiment to ensure there would be enough variance for the main variables (i.e., issue involvement and attitude discrepancy), as well as successful manipulations of the different levels of language intensity. Specifically, 6 issues that are controversial and well-known in the U.S. were selected based on responses from public opinion surveys (ISideWith, 2023), including abortion, immigration, gun control, cryptocurrency, GMO food, and animal testing for medical research. Half of these issues (i.e., abortion, immigration, gun control) received responses that indicate their high issue importance and half received responses indicating lower issue importance in the population. Selecting issues in this manner was intended to ensure there was enough variance in perceived issue involvement, because the LICE model proposes that the effects of language intensity and attitude discrepancy on correction effectiveness will vary between audiences with high and low levels of issue involvement. As hoped, results of the pilot test showed adequate variation in perceived involvement with the 6 issues. Perceived language intensity responses also reflected successful manipulations of language intensity levels in the corrections among the pilot test participants, $t(60) = -9.98$, $p < .001$, Cohen's $d = -2.53$.

In the main experiment, participants were first asked to rate and rank their perceived personal involvement with the 6 issues, and then to rate their prior attitude toward the 6 issues (e.g., anti- or pro-abortion). Demographics including age, sex, race, education, and income were also measured. Next, individuals were randomly assigned to read a misinformation news story on their most or least involved issue as self-reported. For

[2] Attitude discrepancy will technically be tested as a measured rather than manipulated variable.

example, individuals who ranked gun control as their most involved issue and GMO food as their least involved issue saw a misinformation story on either gun control or GMO food. While each issue has a supporting and opposing version of misinformation stories for each issue, the one misinformation story presented to an individual was always pro-attitudinal (i.e., it supported their pre-existing opinion about the controversial issue). This was accomplished by automatically assigning the attitude-consistent version (either support or oppose) based on their prior attitude rating. For example, when the most involved issue was assigned (e.g., gun control), individuals read misinformation promoting the policy if they expressed support for gun control in their prior attitude responses; conversely, misinformation opposing the policy was presented if they indicated they were against gun control in their prior attitude responses. This was done to make people more likely to believe the misinformation due to motivated reasoning, so that the correction (which corrects the misinformation) was always on the opposite side of the misinformation verdict, and thus was counter-attitudinal to all participants when they believed the misinformation.

After reading the misinformation story, participants were asked to rate their belief in the veracity of the misinformation, which was used to indicate their attitude discrepancy with the following correction (i.e., the discrepancy between the correction's verdict and the individual's belief in the veracity of the message being corrected). In this way, most of the individuals in the experiment would have at least some attitude discrepancy with the correction, either low or high depending on the strength of their beliefs in the veracity of the misinformation. This means that individuals who strongly believed the misinformation would have a high attitude discrepancy with the correction; individuals who slightly believed the misinformation would have a low attitude discrepancy with the correction. Individuals who

disbelieved the misinformation had no attitude discrepancy with the correction. See pages 62 and 63 for more detail on the operationalization of attitude discrepancy.

Next, participants read a correction of the misinformation they just saw, which was randomly selected from one of three sources (i.e., *CNN*, *FOX News*, *Insight News*). *Insight News* is a made-up source that indicates no clue of its partisanship to ensure no prior credibility perceptions exist. FOX and CNN were selected as familiar sources representing different political ideologies, such that conservatives would see a correction from either a high-credibility source (i.e., FOX) or a low-credibility source (i.e., CNN), and vice versa for liberals.[3] This was not done to introduce a new factor in the research design, rather it is necessary to ensure variance in perceived correction source credibility in the version of the LICE model where prior source credibility perceptions exist (Figure 3). To ensure there were relatively equal numbers of participants in the high- and low-credibility group, equal numbers of Democrats and Republicans were recruited using screening criteria from Prolific.

The correction displayed to each participant was either high or low in its language intensity level, as determined in the pilot study and confirmed for participants in the main study, which will be discussed in Chapter 5. Examples of the low and high language intensity corrections are provided in the Materials section below. After reading the correction, participants were asked to rate their perceived credibility of the misinformation that was corrected, which is used to indicate correction effectiveness by comparing this credibility rating to the control group who saw no correction. The prompt for this was: “Now that you’ve read the fact-check, we’d like to ask your thoughts about the original news story

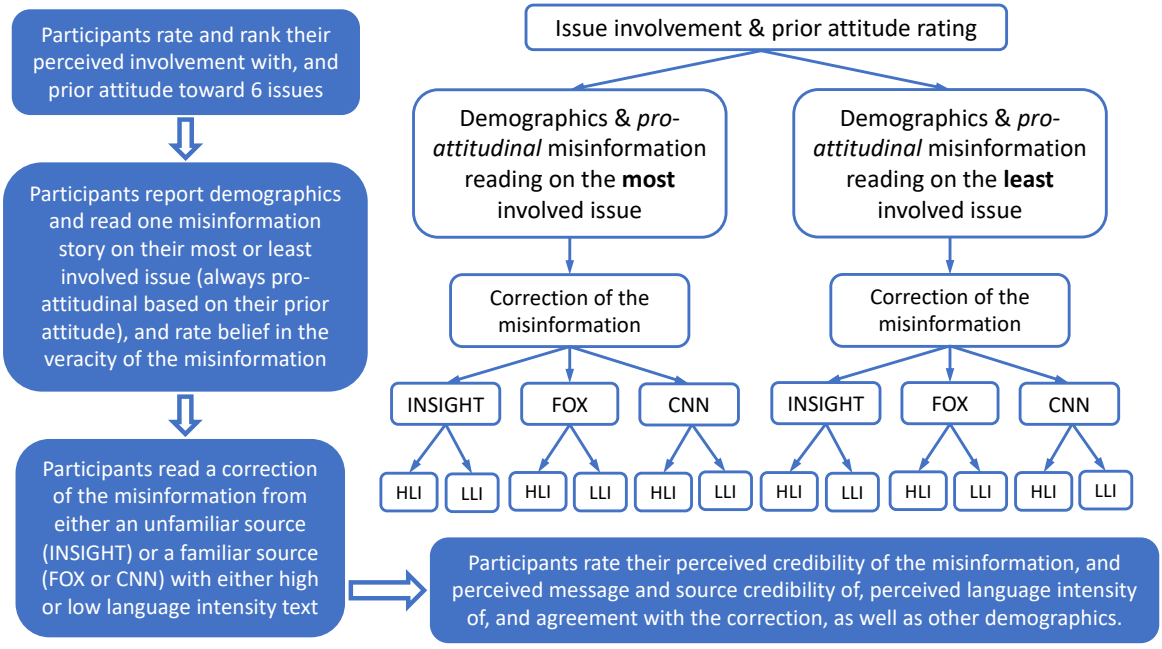
[3] Indeed, analysis of the main study data revealed that conservatives rated FOX as more credible than liberals did, and liberals rated CNN as more credible than conservatives did.

again. To what extent do you feel the original news story you read earlier is:" was used to focus participants to evaluate the original story being corrected in terms of its credibility. Then, with the prompt "The remaining questions on this page ask you to focus back on the fact-check," perceived message credibility, perceived source credibility, and perceived language intensity of the correction were asked in random order. Finally, attention check questions, familiarity with the correction source they saw (as a manipulation check of the familiar vs. unfamiliar source), political party and ideology, social media use and news reading frequency questions were asked, followed by a debrief of the study.[4]

Participants in the control group also rated their issue involvement and prior attitude on the 6 issues, reported demographics, and read an attitude-consistent misinformation story on their most or least involved issue. Then, without seeing a correction, they were asked to rate their perceived credibility of the misinformation right after the misinformation exposure (which provides a measure of baseline misbeliefs), as well as perceived credibility of the three news outlets as fact-checking sources (which provides baseline source credibility perceptions). In the sample of 1509 participants, 381 were exposed to corrections from CNN, 396 saw FOX corrections, 491 saw INSIGHT corrections, and 241 saw no corrections. See Figure 4 for the procedure diagram.

[4] Political party and ideology, as well as social media use and news reading questions, were asked at the end to avoid priming participants about their attitude toward the misinformation and its correction.

Figure 4. Procedure Diagram for the Main Experiment



Note. "HLI" refers to high language intensity and "LLI" refers to low language intensity. There is also a control group where participants rate their perceived credibility of the misinformation and the three news outlets without seeing a correction. Attitude discrepancy is not displayed here to incur additional experimental groups because it is a continuous variable that will be measured instead of being manipulated.

Materials

All news materials used fictional stories to avoid prior exposure to the stimuli. The misinformation stories were created to have approximately the same textual length, tone, and plausibility for each supporting and opposing version on each issue. Corrections for each false story were created with two alternative forms (high and low language intensity) by substituting modifiers and verb forms that have been found to connote different levels of language intensity (Hamilton & Hunter, 1998), as well as different text fonts and punctuation (e.g., capitalization, the exclamation mark). Modifiers such as "positively," "greatly," "most," "definitely," and "extremely" are consistently rated as being of high intensity, and modifiers such as "perhaps," "possibly," "some," "slightly," "somewhat" are generally perceived as

being of low intensity. Similarly, verb forms such as “is,” “causes,” and “must” are perceived as more intense than “seems to be,” “may cause,” and “could.” For example, for a misinformation story with the title “Undocumented Immigrants Strain U.S. Welfare System, Costing $113 Billion Annually,” the correction was presented in a high language intensity format as:

> “A recent story posted by USA Today claims that undocumented immigrants strain the U.S. welfare system, costing $113 billion annually.
>
> This story is FALSE! The CIS report definitely did NOT indicate a significant financial burden from undocumented immigrants accessing U.S. public welfare programs. Instead, it only described some general concerns by the public about this issue that are not supported by ANY data.
>
> IN FACT, the U.S. Immigration and Nationality Act (INA) explicitly BANS undocumented immigrants from accessing nearly all benefits, including welfare, food assistance, non-emergency health coverage, disability coverage, and public or assisted housing.
>
> At the same time, economists have proved that undocumented immigrants make substantial contributions to the U.S. economy through their labor force participation, tax payments, and entrepreneurial endeavors.
>
> Data from the U.S. Census Bureau strongly indicates that undocumented immigrants contribute billions of dollars annually in state and local taxes, which greatly outweighs the costs associated with the very few public services they can access. FAKE NEWS like this story should be banned!

Read more at the link in bio."

Or in a low language intensity format as:

"A recent story posted by USA Today claims that undocumented immigrants strain the U.S. welfare system, costing $113 billion annually.

This story contains some inaccuracies. The CIS report did not indicate a significant financial burden from undocumented immigrants accessing U.S. public welfare programs. Instead, it described some general concerns by the public about this issue that are not supported by data.

The U.S. Immigration and Nationality Act (INA) prevents undocumented immigrants from accessing many benefits, including welfare, food assistance, non-emergency health coverage, disability coverage, and public or assisted housing.

At the same time, some economists have suggested that undocumented immigrants make contributions to the U.S. economy through their labor force participation, tax payments, and entrepreneurial endeavors.

Data from the U.S. Census Bureau indicates that undocumented immigrants contribute billions of dollars annually in state and local taxes, which may outweigh the costs associated with the public services they can access.

Read more at the link in bio."

See Appendix A for all textual stimuli used in this research.

Both the misinformation story and the correction were created in the format of a post on Instagram with likes, comments and post time being blurred to avoid social

influence on the experimental outcomes. Instagram was chosen over other social media platforms (e.g., X) due to its higher character limit for captions. Each post features the Instagram logo, the source's name and logo, a related image with a news headline reflecting the source's recognizable Instagram style (e.g., USA Today, FOX, CNN), news text, and standard Instagram elements. See Appendix B for example pictures.

Measurements

Perceived Language Intensity

To measure perceived language intensity, seven items from the 7-point Likert scale developed by Hamilton and Stewart (1993) were used. Participants evaluated the correction on the extent to which it is perceived as intense, strong, extreme, forceful, emotional, vivid, and assertive. An average score was calculated, and higher scores reflected more intense language. Cronbach's alpha is .87. The manipulation was checked, and results are reported in Chapter 5.

Perceived Message and Source Credibility of Corrections

Perceived *message credibility* of the correction was measured based on recommendations by Appelman and Sundar (2016) using four items. Participants indicated the extent that they felt the fact-check message they read is accurate, believable, authentic, and credible on a scale ranging from 1 = strongly disagree to 7 = strongly agree. An average score of the four items represents perceived message credibility of the correction. Cronbach's alpha is .97. A five-item scale was used to measure perceived *source credibility* of the correction. Items are based on previously validated source credibility scales including how credible, qualified, trustworthy, objective, and out of goodwill individuals feel the source of

the correction is on a 7-point Likert scale. These are dimensions particularly relevant to credibility of a correction source (Liu et al., 2023). Cronbach's alpha is .94.

Perceived Credibility of the Misinformation

The scale for perceived credibility of the misinformation was also based on the message credibility scale by Appelman and Sundar (2016), which includes 4 items: (a) the news story is believable; (b) the news story is accurate; (c) the news story is credible; (d) the news story is authentic measured on a 7-point Likert scale. Four filler items were also included that assessed general attitudes toward the news story (i.e., well-written, clear, interesting, concise) to prevent participants, particularly those in the control group, from assuming the news story was fake. Cronbach's alpha is .93.

Prior Attitude

Prior attitudes about participants' support or opposition toward the 6 issues were measured on a 7-point Likert scale with the question "To what extent do you agree or disagree with the following statements about the 6 issues? – [e.g., I support abortion]."

Attitude Discrepancy

Attitude discrepancy (i.e., the degree to which the correction confirmed or disconfirmed their belief about whether the story they saw was true) was indexed from individuals' belief in the veracity of the misinformation through two items that were asked about the news story prior to exposure to the correction: "The information in this story seems like it is true" and "This story feels like a real news story (vs. fake news)," each rated on a 1

to 9 scale ($M = 6.93$, $SD = 1.68$).[5] Attitude discrepancy was not coded as a categorical variable, but a continuous variable corresponding to the mean of these two items. Specifically, participants who strongly believed the misinformation was true indicated a "high" discrepancy between their initial veracity beliefs in the misinformation and the correction. Those who believed in the misinformation to a lesser extent reflected a relatively "low" attitude discrepancy with the correction. And individuals who disbelieved the misinformation (i.e., with an average rating score equal to or below 5) showed "no" attitude discrepancy with the correction (16.9% of the sample), and were coded as negative values by subtracting 5 from their raw score.

Issue Involvement

Participants' involvement with the six issues was measured using 4 items based on existing scales (Banerjee & McKeage, 1994; Flora & Maibach, 1990). For example, to assess involvement with the issue of abortion, questions were asked "To what extent do you agree or disagree with the following statements about the issue of … [e.g., abortion]?" with the 4 statements—"The issue of abortion is important to me," "I am interested in the issue of abortion," "I spend time thinking about the issue of abortion," and "The issue of abortion is personally relevant to me" on a 7-point Likert scale. The means for each issue are as follows: abortion (5.37), gun control (5.42), immigration (5.00), GMO food (4.17), cryptocurrency (3.07), animal testing for medical research (4.35). Cronbach's alpha = .92. Participants used the scale to rate their perceived involvement with each issue and then also ranked the 6 issues

[5] Using a 1-9 instead of a 1-7 scale is to capture more nuance in the positive (or agreement) side of the scale and ensure more variance in the variable of attitude discrepancy. The correlation between the two items was significant, $r = .64$, $p < .001$.

in order of importance to them (1 = most important, 6 = least important). Abortion was ranked as the most important issue by most people in the sample (39.8%), and 60.4% of the sample ranked cryptocurrency as the least important issue to them. While the ranking was used to determine which issue the misinformation story focused on that each participant would read (to ensure each participant saw an issue that was either high or low involvement for them personally), their assessment of perceived involvement with the specific issue they read on the 4 items described above was used in the data analysis.

Demographics and Other Variables

Demographic variables were collected from the sample including age, sex, race, education, income, political ideology, social media use and news reading frequency. *Familiarity with the correction source* was assessed by asking participants "To what extent are you familiar with the fact-check source [CNN/FOX/Insight News] you saw today?" from 1 (not at all) to 5 (extremely). This variable was used as a manipulation check for the familiar vs. unfamiliar source distinction. *Using social media for news frequency* was reported by asking "How often do you use social media to get news?" from 1 (Never) to 5 (Always). *News reading frequency* was asked with "How many days per week do you watch or read the news from any source?" from 1 (less than one day per week) to 7 days. The frequency variables were used as potential control variables in the data analysis. See full instrument used in this study in the Appendix C.

Sample Characteristics

Participants' average age was 40.65 years (*SD* = 13.33), 66.2% participants were female and 32.6% were male (1.2% did not report their sex or were non-binary). 68.7% of the sample self-identified as white, 14.5% as African American, 6.6% as Spanish, Hispanic,

or Latino, 6.4% as Asian, 0.6% as American Indian, 0.1% as Native Hawaiian or other pacific islanders, and 3.2% as mixed race or preferred not to say. Their average income was between $50,000-$75,000. 12.3% completed high school diploma or less, 34.4% completed some college or associates, 34.2% completed a Bachelor's degree, and 19.0% earned graduate or professional degree (0.1% preferred not to say). The sample contained the approximately same proportion of Republicans (47.1%) and Democrats (48.5%). An additional 3.4% self-identified as Independent, and 1.1% as "other," no preference, or missing.[6] Political ideology was measured by asking individuals to place themselves on a scale ranging from 1 (extremely liberal) to 7 (extremely conservative), and the mean was 3.80 (*SD* = 1.97). The sample's average frequency of using social media for news was between "sometimes" (32.1%) to "most of the time" (30.2%) (*M* = 3.00, *SD* = 1.18). The average news reading frequency was 5.64 days per week (*SD* = 2.27). Statistical analyses showed that participants' demographics were similar across conditions.

[6] Although only Republicans and Democrats were recruited using Prolific pre-screening criteria, some individuals self-reported as Independents or other during the survey and were still included in the data analysis, because they were still able to perceive the correction's source as either having high or low credibility.

Chapter 5 Results

The results of the main study data, along with the validity checks for the key variables, are discussed below.

Manipulation Checks

Manipulations of the independent variables were first checked before analyzing the data. To test if language intensity was successfully manipulated by the text modifications, a t-test was conducted to compare perceived language intensity between individuals in the high and low language intensity conditions. Results showed an overall significant difference in perceived language intensity of corrections between individuals in the low language intensity condition ($M = 3.83$, $SD = .97$) and those in the high intensity condition ($M = 4.98$, $SD = .96$), $t(1260) = -21.19$, $p < .001$, Cohen's $d = -1.19$. The difference test results for each issue are displayed in Table 1. The manipulation of familiar vs. unfamiliar source was checked by comparing individuals' familiarity with the different correction sources. *CNN* ($M = 3.18$, $SD = 1.16$) was perceived as significantly more familiar than *Insight News* ($M = 1.25$, $SD = .70$), $t(863) = 30.08$, $p < .001$, Cohen's $d = 2.06$. Participants were also significantly more familiar with *FOX* ($M = 3.26$, $SD = 1.22$) than *Insight News*, $t(877) = 30.53$, $p < .001$, Cohen's $d = 2.07$. To test if participants were involved with the 6 selected issues in different levels, a repeated measures ANOVA was conducted, and perceived involvement with GMO, cryptocurrency, and animal testing were significantly lower than the other 3 issues ($p < .001$).[7]

[7] This test is to ensure the validity of ranking in issue involvement in case participants rated all high or all low in their involvement levels across all issues.

Table 1. The Difference in Perceived Language Intensity between Low and High Language Intensity Condition for Each Issue

Issue	Low (*M*/*SD*)	High (*M*/*SD*)	*t* (df)	Sig.	Cohen's *d*
Gun control	3.98 (.90)	5.05 (.99)	-8.33 (216)	<.001	-1.13
Immigration	3.90 (1.05)	4.82 (.97)	-6.11 (179)	<.001	-.91
Abortion	3.77 (.96)	5.13 (.89)	-10.67 (213)	<.001	-1.46
GMO	3.88 (1.00)	5.02 (.95)	-8.43 (208)	<.001	-1.16
Animal testing	3.96 (.95)	4.87 (.99)	-6.39 (183)	<.001	-.94
Cryptocurrency	3.56 (.91)	4.96 (.96)	-11.92 (251)	<.001	-1.50

In addition, to test if story version affects the outcome variable, perceived credibility of the misinformation story among the 12 story versions was checked via ANOVAs separately for participants in the control group and experimental groups. For the control group, results showed a significant difference between the pro-gun control story ($M = 5.83$, $SD = .75$) and the pro-cryptocurrency story ($M = 4.42$, $SD = 1.29$) in their credibility perceptions ($p = .02$), as well as between the pro-animal testing story ($M = 5.90$, $SD = 1.01$) and the pro-cryptocurrency story ($p = .003$). All other pairwise comparisons were not significant. For the experimental groups, anti-gun control, pro-gun control, and anti-immigration stories were perceived as more credible than anti-animal testing story ($p < .05$). Pro-cryptocurrency story was perceived as less credible than anti-gun control, pro-gun control, and anti-immigration stories, as well as pro-abortion, pro-GMO, and anti-cryptocurrency stories ($p < .05$). Story version was thus controlled in the hypothesis testing.

Prior Source Credibility Validity Check

To check the validity of treating post-correction source credibility perceptions as prior perceived source credibility in the model for familiar sources, t-tests were conducted between baseline source credibility perceptions of the familiar sources (i.e., perceived source credibility of FOX and CNN in the control group) and those in the experimental groups after

exposure to corrections. Results showed that there were no significant differences between credibility perceptions of CNN with corrections (M = 4.48) and those without corrections (M = 4.23), $t(457) = 1.93$, $p = .06$, Cohen's $d = .16$, as well as for FOX, $t(461) = -.33$, $p = .74$, Cohen's $d = 1.69$. This shows that individuals' existing credibility perceptions of familiar sources were not influenced by a single correction from that source. Thus, although perceived source credibility was asked after exposure to the correction in the familiar source conditions, it can be used to represent individuals' prior credibility perceptions of the correction source.

Correction Effectiveness (Control vs. Experimental)

To test if corrections were generally effective in reducing perceived credibility of the misinformation regardless of correction features, ANCOVAs were conducted to compare the difference in perceived credibility of the misinformation between the experimental groups and the control group. Sex, race, political ideology, and social media use were selected as the control variables based on their significant correlations with the main variables. Story version was also controlled. Results showed a significant difference in perceived credibility of the misinformation between the experimental groups (M = 4.47) and the control group (M = 5.33), $F(1,1483) = 67.03$, $p < .001$, $\eta_p^2 = .04$. This demonstrates that the corrections were generally effective in reducing perceived credibility of the misinformation regardless of their language intensity.

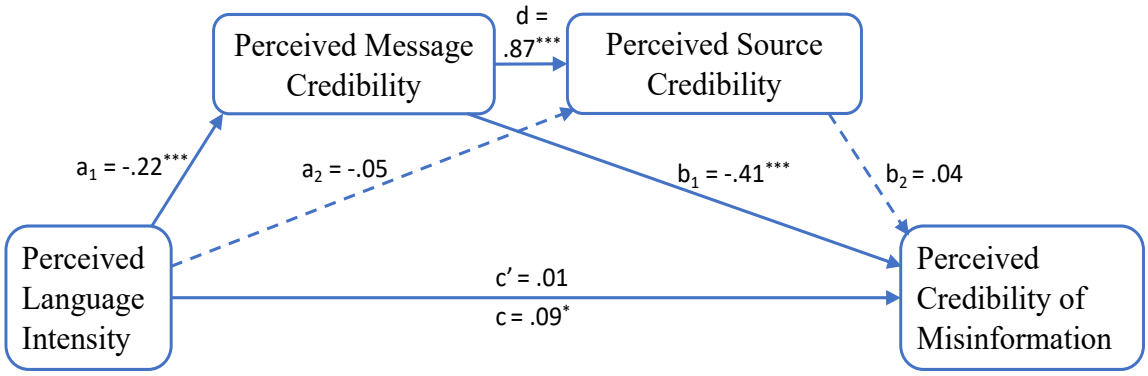
Testing the LICE Model for Unfamiliar Sources

To test the LICE model for correction sources without prior credibility perceptions (Figure 2), a serial mediation model (Model 6) in SPSS PROCESS (Hayes, 2017) was conducted using the data from participants who saw corrections from *Insight News*, while controlling for demographics (i.e., sex, race, political ideology, and social media use),

attitude discrepancy, issue involvement, and story version (n = 489). Individuals with zero or negative values of attitude discrepancy were also included in the analysis to control for variance in all individuals including those who disbelieved the misinformation in the first place.[8] Results showed a significant effect of perceived language intensity on perceived message credibility of the correction (path a_1) (b = -.30, β = -.22, p < .001). Individuals perceived the corrective message with increased language intensity as less credible. Perceived message credibility of the correction also significantly impacted perceived credibility of the misinformation following exposure to the correction (path b_1) (b = -.42, β = -.41, p < .001), such that corrections perceived as more credible were more effective in reducing perceived credibility of the misinformation (i.e., were more successful in correcting misbeliefs). However, the effect of language intensity on perceived source credibility of the correction was not significant (path a_2) (b = -.05, β = -.04, p = .07), nor was the effect of source credibility on misinformation credibility (path b_2) (b = .04, β = .04, p = .66). Message credibility of the correction significantly predicted source credibility of the correction in a positive way (path d) (b = .81, β = .87, p < .001), such that individuals perceived the unfamiliar correction source as more credible when it delivered credible corrections (see Figure 5).

[8] This is to ensure the findings can be generalized to all individuals including those who disbelieved the misinformation initially. The same analyses were also conducted only among individuals who believed the misinformation before correction, and results were consistent in all paths' directions and significance except path a_2 where language intensity significantly and negatively predicted perceived source credibility of the correction, b = -.06, β = -.05, p = .04. This suggests that misinformation *disbelievers* may be more cautious in judging the credibility of both misinformation and correction sources, perhaps because they are more analytical thinkers in general. By contrast, misinformation *believers* are less cautious in drawing conclusions (i.e., they are more likely to easily draw conclusions) about an unfamiliar source based on a correction from the source, and are more easily influenced by the correction's language intensity in making source credibility evaluations.

Figure 5. Model Results for Unfamiliar Sources



Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. Solid lines indicate significant paths and dashed lines indicate non-significant paths. Coefficients displayed are standardized coefficients (β).

When it comes to the mediation processes, the mediation effect of correction message credibility on the relationship between perceived language intensity and perceived credibility of the misinformation was significant (path a_1b_1) (b = .13, β = .09, 95% CI [.04, .15]). H1 proposing that increased perceived language intensity increases perceived credibility of the misinformation by decreasing perceived credibility of the correction message was supported. Source credibility, however, did not mediate the relationship between correction language intensity and perceived post-correction misinformation credibility (path a_2b_2) (b = -.002, β = -.001, 95% CI [-.01, .006]). H3 that proposed perceived source credibility of the correction as a mediator was thus not supported. The serial mediation path was not significant either (path a_1db_2) (b = -.01, β = -.007, 95% CI [-.04, .02]). H2 proposing that increased perceived language intensity reduces perceived message credibility of the correction, which in turn lowers perceived source credibility of the correction was supported, but decreased source credibility did not ultimately lead to reduced correction effectiveness.

Testing the LICE Model for Familiar Sources

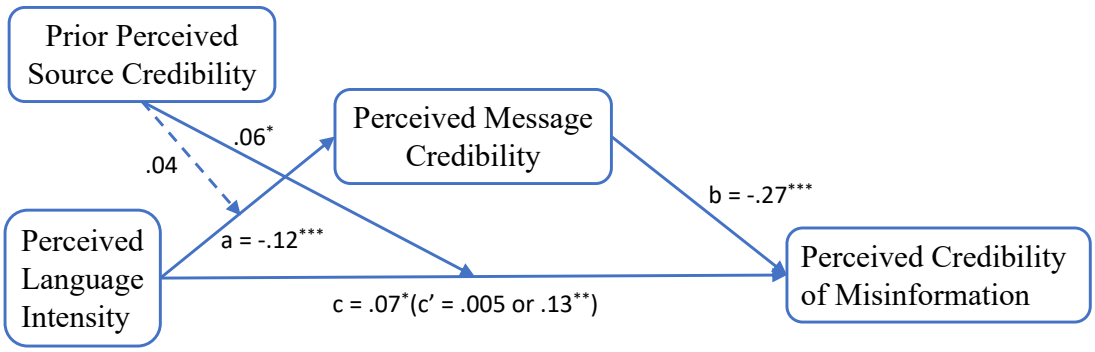
To test the model for correction sources with prior credibility perceptions (Figure 3), a moderated mediation model (Model 8) in PROCESS (Hayes, 2017) was conducted using only the data from participants who saw corrections from FOX and CNN and the control group, while controlling for demographics, attitude discrepancy (including zero and negative values), issue involvement, and story version (n = 761).[9] Results showed a significant effect of perceived language intensity on correction message credibility for familiar sources (path a) (b = -.17, β = -.12, p < .001), as well as a significant path b (i.e., correction message credibility predicted post-correction misinformation credibility) (b = -.26, β = -.27, p < .001). The main effect of correction source credibility on correction message credibility was also significant (b = .77, β = .75, p < .001), such that individuals perceived the corrective message as more credible when it came from a credible source. However, the interaction effect of language intensity and source credibility on correction message credibility was not significant (b = .03, β = .04, p = .07). Thus, H5 was not supported.

Next, perceived language intensity of the correction significantly predicted post-correction misinformation credibility for familiar sources (path c') (b = .09, β = .07, p = .03). Individuals perceived the misinformation as more credible after reading a correction using increased intensity language. The interaction effect of language intensity and source credibility on post-correction misinformation credibility was significant (b = .05, β = .06, p = .03). Simple slopes analysis revealed that when the correction source credibility was low (1

[9] The same analyses were also conducted only among individuals with positive values of attitude discrepancy (i.e., those who believed the misinformation before correction), and results were same in terms of all paths' directions and significance.

SD below the mean), language intensity did not impact perceived credibility of the misinformation (b = .007, β = .005, p = .91). But when source credibility was high (1 SD above the mean), increased language intensity led to greater credibility perceptions of the misinformation (b = .18, β = .13, p = .003). This is contrary to H4 which proposed that increased perceived language intensity leads to reduced misinformation credibility when prior perceived source credibility is high, but increases misinformation credibility when source credibility is low. In other words, when the correction source was perceived as not credible, corrections were less effective in reducing misinformation credibility (M = 4.64 for low source credibility groups) regardless of their language intensity. When the correction source was perceived as credible, corrections with lower language intensity were more effective in reducing misinformation credibility (M = 4.23 for high source credibility x low language intensity group) than those with higher language intensity (M = 4.63 for high source credibility x high language intensity group). See Figure 6 for the model results.

Figure 6. Model Results for Familiar Sources



Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. Solid lines indicate significant paths and dashed lines indicate non-significant paths. Coefficients displayed are standardized coefficients (β). Coefficients for the moderation lines (i.e., .04 and .06*) represent the moderation effect index only, but not simple slopes analysis results. Direct effects of language intensity on misinformation credibility depend on the source credibility level, such that c' = .005 when source credibility is low, and c' = .13** when source credibility is high.

Additionally, the moderated mediation was not significant (b = -.009, β = -.01, 95% CI [-.02, .004]). Although there was a mediation effect of message credibility on the relationship between language intensity and misinformation credibility, it did not depend on the level of source credibility as a moderator. Higher level of the source credibility led to a similar mediation effect (b = .03, β = .02, 95% CI [.02, .07]) as the lower level of source credibility (b = .06, β = .04, 95% CI [.001, .05]).

Testing the LICE Model for Individual Variables

Hypothesis 6 and 7 that proposed attitude discrepancy and issue involvement would moderate the relationship between perceived language intensity and misinformation credibility when keeping source credibility constant were tested using the moderated moderation model (Model 3) in PROCESS (Hayes, 2017) and the data from all experimental conditions, while controlling for perceived correction source credibility, demographics, and story version (n = 1041). Only participants with positive attitude discrepancy (i.e., those who believed the misinformation before seeing the correction) were included, as the effect of attitude discrepancy may be most prominent among those individuals who also need effective corrections the most (i.e., those who believe the misinformation to be true). This additionally helps avoid a floor effect in misinformation credibility among individuals who initially disbelieved the misinformation. Results showed a significant main effect of language intensity of the correction on misinformation credibility (b = .08, β = .06, p = .046), as well as attitude discrepancy on misinformation credibility (b = .39, β = .28, p < .001). The interaction effect of correction language intensity and attitude discrepancy was also significant (b = -.08, β = -.06, p = .02). Simple slopes analysis revealed that when attitude discrepancy was low (1 SD below the mean, indicating relatively weak initial belief the

misinformation was true), increased language intensity used in the correction led to greater credibility perceptions of the misinformation (b = .16, β = .12, p = .005). When attitude discrepancy was high (1 SD above the mean, indicating relatively strong initial belief the misinformation was true), language intensity did not significantly impact misinformation credibility (b = .001, β = .001, p = .97). H6 proposing that increased perceived language intensity reduces correction effectiveness when attitude discrepancy is high, but improves correction effectiveness when attitude discrepancy is low was not supported. In other words, for individuals who only initially believed the misinformation to a small extent, corrections with increased language intensity reduced effectiveness in reducing belief in the misinformation. When individuals initially believed the misinformation to a greater degree (above the mean of 7.62 in a 9-point scale), they were unaffected by the language intensity of the correction in assessing misinformation credibility.

The three-way interaction effect between perceived issue involvement, attitude discrepancy, and perceived language intensity on misinformation credibility turned out not to be significant (b = -.01, β = -.01, p = .70), nor were the interaction effects between issue involvement and attitude discrepancy (b = -.01, β = -.02, p = .52) or involvement and language intensity (b = .01, β = .02, p = .44). Because no moderated moderation was found, H7 proposing that the effect in H6 will be strengthened with increased issue involvement was not supported. In addition, perceived issue involvement significantly and positively predicted perceived credibility of the misinformation when controlling for other variables (b = .06, β = .08, p = .01), such that more issue involvement led to greater credibility perceptions of the misinformation among individuals who believed the misinformation.

Moreover, to explore the effect of individual differences among individuals who disbelieved the misinformation in the first place, the same moderated moderation analysis was conducted among participants with zero or negative attitude discrepancy values only (n = 210). Language intensity (b = .22, β = .16, p = .01) and attitude discrepancy (b = .55, β = .39, $p < .001$) significantly predicted misinformation credibility as well. However, the interaction effect between language intensity and attitude discrepancy was not significant ($b < .001$, $\beta < .001$, p = .99). The effect of language intensity on misinformation credibility did not depend on the level of attitude discrepancy among individuals who disbelieved the misinformation. In other words, corrections with increased language intensity increased misinformation credibility even among individuals who initially disbelieved the misinformation, and this is regardless of the extent of their disbelief. Other two-way and three-way interactions were not significant either, although issue involvement also positively predicted misinformation credibility for these individuals (b = .17, β = .22, p = .01).

Combined Effects of Message, Source, and Audience Features

To explore the combined effects of perceived language intensity, perceived source credibility, attitude discrepancy, and perceived issue involvement, multiple regressions were conducted by entering the demographics and story version in Block 1, the four independent variables (mean-centered) in Block 2, the three two-way interactions in Block 3 (i.e., language x source/attitude/involvement), the three three-way interactions in Block 4 (i.e., language x source x attitude/involvement, language x attitude x involvement), and the one four-way interaction in Block 5 (n = 1041).[10]

---

[10] Given the central importance of language intensity in the model, only interactions between language intensity and other features were explored, rather than all possible interactions. Only individuals with positive values of

Results of the final model showed significant main effects of sex (female = 0, male = 1), correction source credibility, attitude discrepancy, and issue involvement on misinformation credibility, but not for language intensity (see Table 2). There was a significant interaction effect between language intensity and attitude discrepancy (b = -.08, β = -.06, p = .03), and the three-way interaction between language intensity, source credibility, and issue involvement was also significant (b = .03, β = .09, p = .005). All other interactions were not significant. Simple slopes analysis revealed that when perceived correction source credibility and issue involvement were both low, increased language intensity led to greater credibility perceptions of the misinformation (b = .18, β = .13, p = .007). When perceived correction source credibility and issue involvement were both high, increased language intensity also increased misinformation credibility (b = .28, β = .20, p < .001), indicating that corrections with high-intensity language are less effective in reducing misinformation credibility than corrections that employ less intensive language. Language intensity of the correction had no effect on misinformation credibility in other combinations of correction source credibility and issue involvement. See Figure 7 and Figure 8 for the effect visualization.

attitude discrepancy (i.e., those who believed the misinformation before correction) were included in the attitude discrepancy variable. The same analyses were also conducted only among individuals who disbelieved the misinformation (n = 210), and no interaction effects were significant. This suggests that those individuals were less influenced by the correction and its various features in altering misinformation credibility, emphasizing more practical use on targeting individuals who are most vulnerable to misinformation.

Table 2. Regression Analysis of Interactions between Perceived Language Intensity, Source Credibility, Attitude Discrepancy, and Issue Involvement Predicting Perceived Credibility of Misinformation

Predictor	Model 1		Model 2		Model 3		Model 4		Model 5	
	b (SE)	*β*	*b (SE)*	*β*	*b (SE)*	*β*	*b (SE)*	*β*	*b (SE)*	*β*
Sex	.29 (.10)	.09**	.24 (.09)	.08**	.24 (.09)	.08**	.24 (.09)	.08**	.24 (.09)	.08**
Race	.34 (.10)	.11***	.10 (.09)	.03	.11 (.09)	.03	.12 (.09)	.08	.12 (.09)	.04
Political ideology	.02 (.02)	.02	.04 (.02)	.06*	.04 (.02)	.05	.04 (.02)	.05	.04 (.02)	.05
Social media use	.03 (.04)	.02	.06 (.04)	.05	.06 (.04)	.04	.06 (.04)	.04	.06 (.04)	.04
Story version	-.07 (.01)	-.16***	.002 (.02)	.004	.002 (.02)	.005	.001 (.02)	.002	.001 (.02)	.003
Language intensity			.06 (.04)	.05	.07 (.04)	.06	.07 (.04)	.06	.07 (.04)	.05
Source credibility			-.32 (.03)	-.33***	-.32 (.03)	-.34***	-.32 (.03)	-.34***	-.33 (.03)	-.34** *
Attitude discrepancy			.40 (.04)	.29***	.39 (.04)	.29***	.38 (.04)	.28***	.38 (.04)	.28***
Issue involvement			.06 (.03)	.08*	.06 (.03)	.09*	.08 (.03)	.11**	.08 (.03)	.11**
Language x source					.02 (.02)	.02	.01 (.02)	.02	.02 (.02)	.03
Language x attitude					-.08 (.03)	-.06*	-.08 (.03)	-.06*	-.08 (.03)	-.06*
Language x involvement					.01 (.02)	.02	.02 (.02)	.03	.02 (.02)	.03
Language x source x attitude							-.002 (.02)	-.003	-.004 (.02)	-.01
Language x source x involvement							.03 (.01)	.08**	.03 (.01)	.09**
Language x attitude x involvement							-.001 (.02)	-.003	-.003 (.02)	-.01
Language x source x attitude x involvement									-.01 (.01)	-.03
Adjusted R^2	.04		.26		.26		.26		.26	
ΔR^2	.05***		.22***		.004		.006		.001	
F	10.35		41.20		31.50		25.84		24.26	
Sig	<.001		<.001		<.001		<.001		<.001	

Notes: *$p < .05$, **$p < .01$, ***$p < .001$

Figure 7. The Effect of Perceived Language Intensity on Misinformation Credibility at Different Levels of Issue Involvement When Source Credibility is Low

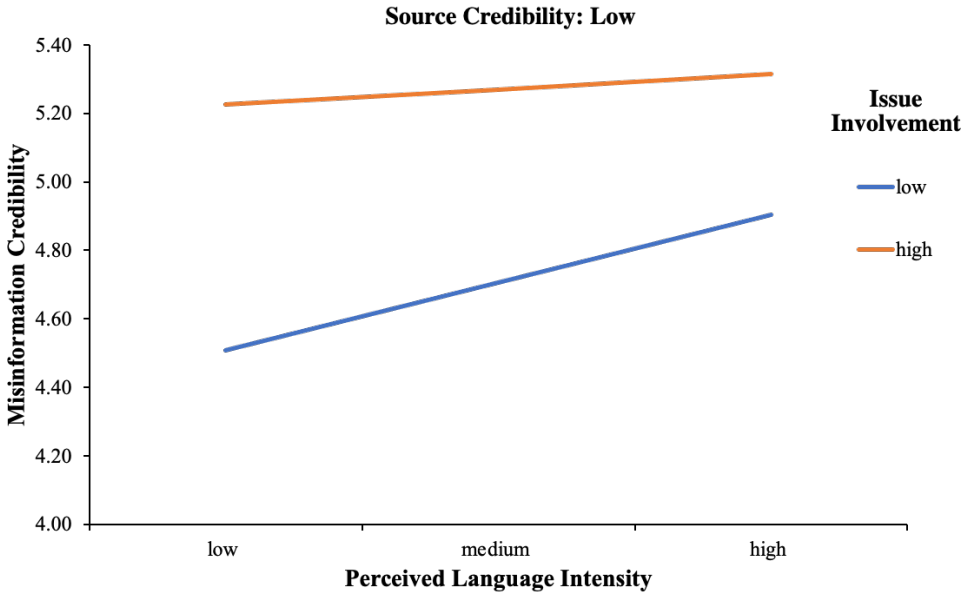
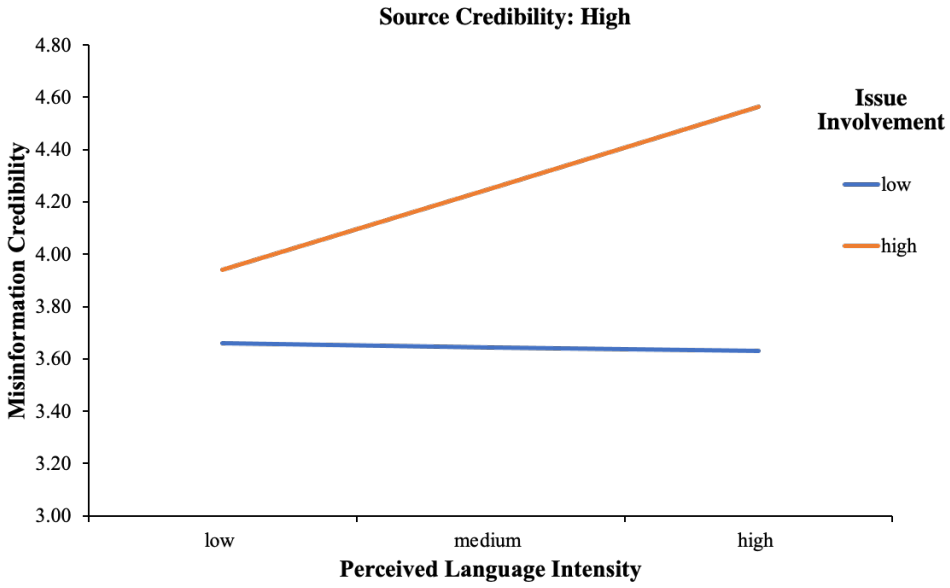


Figure 8. The Effect of Perceived Language Intensity on Misinformation Credibility at Different Levels of Issue Involvement When Source Credibility is High



Chapter 6 Discussion

This dissertation study aims to address a significant gap in the misinformation literature by taking message, source, and audience factors into account to evaluate correction effectiveness. Specifically, a model of language intensity and correction effectiveness (i.e., the "LICE" model) was proposed and tested to investigate perceived language intensity of the correction as a central message feature, and its effect on correction effectiveness under the influence of other source and individual features including perceived source credibility of the correction, attitude discrepancy, and perceived issue involvement. Moreover, effects of correction source credibility were discussed in two contexts—familiar source with prior credibility perceptions, and unfamiliar source without existing credibility perceptions. Overall, results revealed general effectiveness of corrections in reducing misinformation credibility compared to the no-correction control group, regardless of difference in language intensity levels and other features.

Effects of Source and Message Features

Findings for both familiar and unfamiliar sources confirmed the prediction that increasing language intensity leads to reduced correction effectiveness (i.e., increased misinformation credibility), because language intensity reduced perceived message credibility of the correction. This suggests that even in the misinformation context, reactance theory and the language expectancy theory (LET) apply in a similar manner as they do in the health campaign context. In both cases, high-intensity language may activate resistance and/or violate the audience's professional and normative expectations for correction messages. In any case, future research should delve deeper into verifying the role of reactance and expectancy violation processes in the misinformation correction context. While

some assertiveness is recommended for fact-checks to convey certainty and authority (Ferracioli et al., 2022), the high-intensity language used in this study demonstrated the opposite effect, backfiring in its ability to reduce belief in misinformation. This underscores the need for a delicate balance between assertiveness and intensity: fact-checkers must project confidence and clarity without overwhelming the audience with language that is perceived as too forceful or aggressive.

Additionally, this research calls for a clearer definition and operationalization of language intensity within the specific context of news and misinformation correction. Without a clear, shared understanding of what constitutes "intense" language, researchers and fact-checkers might apply or evaluate correction strategies inconsistently, leading to mixed results in misinformation correction efforts. A precise operationalization could also help determine the optimal level of intensity needed to assert facts without triggering negative reactions. Identifying this fine line between being assertive and overly intense is essential to ensuring that corrections are persuasive without diminishing their credibility or effectiveness. Recommendations for future research along these lines are offered later in this chapter.

For unfamiliar sources, the LICE model predicted that individuals form credibility impressions of the source based on their evaluations of the message it delivers. This concept of "credibility transfer" (Schweiger, 2000) suggests that when people assess the credibility of a message, their judgment can extend to influence how credible they perceive the source itself to be. In support of this, the findings revealed that perceived credibility of the correction message positively predicted the credibility of the unfamiliar source. However, contrary to expectations, correction source credibility did not significantly influence misinformation credibility, nor was source credibility affected by language intensity of the

correction. This suggests a nuanced process where individuals evaluate a correction message's credibility based on its linguistic features, and while they may transfer these credibility judgments to the correction source, this transfer does not extend to altering how they perceive the credibility of the misinformation itself. Thus, while message features can shape source credibility indirectly by shaping message credibility first, they appear to have limited reach in shaping beliefs about the misinformation being corrected through source credibility only. The important role of source credibility on correction effectiveness suggested in the literature is therefore diminished when dealing with unfamiliar sources. In such cases, correction message credibility becomes the more crucial factor in determining the success of corrections in reducing belief in misinformation, highlighting the significance of considering message features during correction efforts that are from sources people are less familiar with.

For familiar sources, the LICE model predicted that when (prior) source credibility is high, increased language intensity enhances both message credibility and correction effectiveness. Conversely, when source credibility is low, heightened language intensity has the opposite effect, reducing message credibility and correction effectiveness. These predictions align with the LET, which posits that high-credibility sources may have a broader tolerance for using intense language without negatively violating audience expectations. The information processing theory of language intensity effects (Hamilton et al., 1990; Hunter et al., 1984) also suggested greater attitude change with the combination of high-intensity language and high source credibility.

However, despite suggestions from these theoretical frameworks and supportive evidence from persuasion research (e.g., Buller et al., 2000; Burgoon et al., 1975), the

findings in this study tell a different a story. Rather than increasing credibility, high-intensity language reduced perceived message credibility of the correction regardless of whether the source had high or low credibility. Corrections delivered with intense language were also less effective at reducing belief in misinformation than those with less intensity, even when they came from high-credibility sources. This suggests that, while high-credibility sources may generally be more persuasive, the use of intense language can undermine the source's persuasive power, by possibly leading to skepticism or resistance from the audience. When the correction source was perceived as lacking credibility, corrections were generally ineffective at reducing belief in misinformation, irrespective of the language intensity used. Therefore, for low-credibility sources, neither the message nor the manner of its delivery significantly influences belief change. These findings challenge the expectation that source credibility can buffer against the negative effects of intense language and highlight the impact of language intensity in the misinformation correction context, even for familiar, trusted sources.

The LET and the information processing theory of language intensity effects were largely developed in contexts such as health or persuasive communication, where the stakes and expectations of the audience differ from the context studied in this dissertation. It is likely that in the misinformation context individuals have heightened sensitivity to the language used in corrections, making them more skeptical of intense language, even from trusted sources. Intense language may come across as overly emotional or biased that can reduce trust in the message, particularly when objectivity and neutrality are expected in correction efforts, which would be less expected in health campaign messages. Thus, it is

crucial to consider the specific context when applying LET and related frameworks, as the impact of language intensity can vary significantly depending on the situation.

Overall, these results reveal a distinct process in the misinformation correction context, where individuals critically examine cues to assess credibility, even when encountering pro-attitudinal information. First, they tend to evaluate the source of the correction. If the source is unfamiliar and no prior credibility perceptions exist, individuals shift their focus to the message content and delivery style. In this case, high-intensity language serves as a discounting cue, indicating low message credibility, which in turn lowers perceived credibility of the source and correction effectiveness in reducing belief in misinformation. For familiar sources, the process shifts slightly. When individuals have pre-existing low-credibility perceptions of a familiar correction source, the correction tends to be ineffective, as strong negative perceptions of the source override any potential influence of the message cues. On the other hand, when the correction source is perceived as credible, individuals still look for cues within the message itself. If intense language is used, it undermines the message's credibility and effectiveness, and the high credibility of the source is not enough to counterbalance the negative impact of the language. Therefore, for a correction to be truly effective, both source cues and message cues must align in credibility. If either the source or the message is perceived as lacking credibility, the overall effectiveness of the correction in reducing belief in misinformation is diminished. This highlights the importance of a balanced approach, where both the source and the message content must be carefully managed to maintain trust and credibility when it comes to correcting the misinformation.

Effects of Individual and Message Features

To investigate the effects of individual features, the LICE model predicted that, when controlling for source credibility, increased language intensity reduces correction effectiveness when attitude discrepancy is high, but improves effectiveness when attitude discrepancy is low. This prediction was based on the nonlinear information processing model (Hamilton & Stewart, 1993), which posits that the pairing of high language intensity and high attitude discrepancy leads to greater attitude change until discrepancy increases to a critical point, beyond which high intensity will inhibit attitude change.

In contrast to these expectations, the findings revealed a negative impact of high-intensity language on correction effectiveness among individuals with low attitude discrepancy. For these individuals, who only slightly believed the misinformation and may harbor at least some doubt about its veracity, the use of intense language in corrections appears to diminish their existing doubts on the misinformation. As a result, they perceived the misinformation as more credible compared to those exposed to corrections with less intense language, although still less credible than the control group, as determined through an additional t-test, $t(773) = 11.06$, $p < .001$, Cohen's $d = .86$. In other words, rather than reinforcing skepticism toward the misinformation, high-intensity language diminishes the correction's efficacy in reducing perceived credibility of the misinformation among individuals with low attitude discrepancy. This suggests that the tolerance for using high-intensity language in misinformation corrections may be extremely low compared to in health or environmental campaign messages, leading to reduced effectiveness even among individuals who held only slight attitude discrepancy with the correction. Therefore, carefully considering language intensity in corrections is essential when targeting individuals with

minimal initial belief in the misinformation, as they may be more influenced by the message feature and thus receptive to belief adjustment.

On the other hand, individuals who strongly believed the misinformation (i.e., those with high attitude discrepancy with the correction) were relatively unaffected by language intensity of the correction. In this scenario, regardless of how intense the language may be, they reported higher perceived credibility of the misinformation than those with low attitude discrepancy, although still lower than the control group as shown by a t-test, $t(761) = 2.10$, $p = .02$, Cohen's $d = .16$. This result may stem from how attitude discrepancy was operationalized—based on an average belief score in the misinformation's veracity (see pages 78 & 79)—which closely mirrors perceived misinformation credibility. Thus, it is unsurprising that more attitude discrepancy led to more misinformation credibility after exposure to the correction. High baseline beliefs in the misinformation's veracity likely accounted for higher post-correction misinformation credibility, even if their misbeliefs shifted to some extent.

To address this, additional analyses were conducted to explore perceived message credibility of the correction, rather than misinformation credibility, as an indicator of individuals' reception to the correction. Results showed that attitude discrepancy negatively predicted message credibility.[11] Individuals with high attitude discrepancy found the correction message less credible than those with low discrepancy, which aligns with the

[11] The same process model was used to test attitude discrepancy as a moderator in the relationship between language intensity and message credibility of the correction with same covariates. There was a significant main effect of attitude discrepancy on message credibility, $b = -.06$, $\beta = -.04$, $p = .02$, but not an interaction effect between language intensity and attitude discrepancy, $b = -.02$, $\beta = -.02$, $p = .30$.

confirmation bias and motivated reasoning theory (Kunda, 1990). Considering the significant positive effect of message credibility on correction effectiveness, it is thus likely that corrections were indeed less effective for those with strong prior misbeliefs, despite the intensity of language. This indicates that once individuals are deeply entrenched in a belief, the effectiveness of corrections is diminished, irrespective of the emotional or persuasive power of the language used. In conclusion, taken together, the results suggest that corrections may be most effective for individuals with uncertain or alterable prior beliefs in the misinformation, but also caution that using corrections with high-intensity language may inadvertently strengthen misbeliefs for this group. For those who strongly believed in the misinformation and distrust the correction, other correction strategies, rather than altering language features, are needed to reduce misbeliefs, such as using in-group or highly credible correction sources or employing narrative formats in a correction.

Moreover, the negative effect of high-intensity language on correction effectiveness extended to individuals who initially disbelieved the misinformation (i.e., those with no attitude discrepancy with the correction). They also perceived misinformation as more credible when exposed to high-intensity corrections compared to low-intensity ones, although they still perceived the misinformation as less credible than the control group.[12] Therefore, even when initial beliefs aligned with the correction, high-intensity language undermined correction effectiveness, indicating a broader detrimental effect of intense language on correction outcomes.

---

[12] An additional t-test was run to test the difference in misinformation credibility between the control group and the experimental group (only among those with no attitude discrepancy), $t(451) = 13.45, p < .001$, Cohen's $d = 1.27$.

Turning now to the other individual variable investigated in this research, although perceived issue involvement was not found to strengthen the effect of language intensity and attitude discrepancy on correction effectiveness, it did significantly predict misinformation credibility. More involved individuals perceived the misinformation as more credible than did less involved individuals, irrespective of language intensity or attitude discrepancy. This finding is consistent with prior research suggesting that issue involvement heightens compliance with issue-promoting messages (A. Clark, 1993; T. Clark, 1998). The fact that a combination of high language intensity, low attitude discrepancy, and high issue involvement did not enhance correction effectiveness as predicted by the LET further confirms the overwhelming negative effect of high-intensity language on correction effectiveness. Because individuals who hold lower levels of misbelief were not more tolerant of the use of high-intensity language and reduced their misbeliefs to a smaller extent than those exposed to less intense corrections, high involvement with the issue could only amplify the negative effect but not mitigate it. This holds true even for individuals who initially disbelieved the misinformation where they were dissuaded by corrections with high-intensity language, and the level of issue involvement could not alter this negative impact.

In summary, the data contradict the initial hypothesis that high-intensity language improves correction effectiveness when attitude discrepancy is low and diminishes it when discrepancy is high. Instead, high-intensity language consistently undermines correction effectiveness across different levels of attitude discrepancy and issue involvement, demonstrating the intolerance for using intense language in misinformation corrections among the audience under any circumstances. Corrections using high-intensity language may even reinforce misinformation credibility, especially among those with low discrepancy and

high involvement, and alternative correction strategies should be considered for individuals with high discrepancy or strong prior misbeliefs.

Combined Effects of Source, Individual, and Message Features

Efforts to explore the combined effects of all features revealed an interesting interaction between language intensity, source credibility, and issue involvement. Corrections with increased language intensity led to increased misinformation credibility, but only when both source credibility and issue involvement were either high or low. Potential reasons for these results are offered below.

When correction source credibility is high, individuals are more likely to trust the overall content of the correction. However, because they are highly involved with the issue, they will also deliberately look for additional available cues, such as language intensity, before making a final judgment. As they devote more cognitive effort to evaluating both the content and features of the message, excessive or forceful language can undermine even a credible source's persuasive power, causing the correction to lose credibility and inadvertently increasing the perceived credibility of the misinformation. Another possible explanation is that individuals may hold high expectations for sources they perceive as highly credible, particularly on issues they are highly involved with. Consequently, intense language used in the correction can violate their expectations easily (e.g., they might be surprised that a high-credibility source they trust would resort to such intense language that seems too unprofessional and biased), and led to reduced correction effectiveness.

On the other hand, when source credibility is low, individuals are less likely to trust the correction message outright, and might ignore the message entirely because they already distrust its source. However, when they are also low in issue involvement, they may lack

prior knowledge about the issue too, leaving them less confident in evaluating the correction solely based on the source. As a result, they may rely on other cues, such as language intensity, to aid their judgment. In this case, intense language combined with a low-credibility source can amplify the negative impact on correction effectiveness, even when individuals felt less certain or knowledgeable about the issue at hand. Future research could deepen our understanding of these dynamics by examining the role of prior knowledge in conjunction with issue involvement on correction effectiveness. Similarly, applying LET to this situation, it is also possible that individuals expect a low-credibility source to communicate in an unprofessional or biased manner. Intense language used by such a source only confirms their expectations, amplifying the negative effect of low source credibility on correction effectiveness. And this only happens when issue involvement is also low, because low involvement provides little motivation for recipients to counterbalance these negative impressions by engaging with the message's underlying arguments, compounding the negative effect of language intensity.

When there is a mismatch between source credibility and issue involvement, the impact of language intensity tends to diminish. For instance, if the source of the correction is credible but the audience is not highly involved with the issue, people may generally trust the message but won't scrutinize its details as carefully. In this scenario, intense language might have little effect because people are relying more on the source's trustworthiness than on specific message characteristics. Conversely, when source credibility is low but issue involvement is high, individuals are motivated to carefully examine the correction because the issue matters to them. However, their distrust of the source may lead them to engage in skeptical processing, where they scrutinize all aspects of the correction, including language

intensity, in a way that confirms their negative evaluation of the source. In this case, language intensity alone doesn't significantly impact correction effectiveness because individuals have already formed a strong negative judgment about the source. Their skepticism outweighs any effect the message tone might have. In both mismatched cases, language intensity has no significant effect because people's judgments are primarily guided by either their distrust of the correction source or their disinterest in the issue, not by the message features themselves.

In conclusion, the combined effects of source, individual, and message features reveal how each component uniquely and interactively contribute to the success of misinformation corrections. Message features can either enhance or hinder persuasion depending on how they interact with other factors, although high-intensity language has the potential to always diminish correction effectiveness in the misinformation context. Source credibility remains a fundamental factor, as people tend to trust or dismiss corrections based on their perception of the source's credibility. Individual factors, such as attitude discrepancy and issue involvement, influence how willingly people accept the correction; those with low discrepancy and low issue involvement are more likely to accept the correction that contradicts their prior beliefs. When all three—message features, source credibility, and individual characteristics—interact in certain ways, they shape whether the correction is effective or not, and reveal distinct strategies for different audience groups. Understanding how these factors combine provides key insights into the complexity of corrective messaging and can guide the development of more targeted and effective strategies for combating misinformation.

Limitations and Future Directions

This dissertation contributes valuable insights into the role of message, source, and individual features in misinformation correction, but several limitations remain. First, while the findings suggest that high-intensity language consistently undermines correction effectiveness, these results are limited to the materials and levels of intensity employed in the study. Future studies should aim to refine the definition and operationalization of language intensity, particularly in (and specific to) diverse communication platforms, in order to identify the optimal level of intensity. This could involve testing different degrees of language intensity across varying levels of source credibility and individual features to determine when intense language enhances correction effectiveness without triggering resistance or skepticism.

Another limitation lies in the generalizability of the findings to real-world settings. The study primarily focuses on the effect of language intensity in a controlled experimental setting. While this allows for precise manipulation and measurement of variables, it may not fully capture the nuances of real-world misinformation correction, where diverse communication platforms and uncontrolled factors (e.g., social influence, media exposure) can influence outcomes. Future research could benefit from conducting field experiments that assess correction effectiveness in naturalistic settings, such as social media platforms or news outlets.

The study design, in which individuals were conditioned to read only pro-attitudinal misinformation, also limits the scope of this research. While this approach reflects a maximized scenario where misinformation is most likely to be believed and shared, and thus where corrections are most needed, it does not encompass other situations. Future research

should explore the effects of language intensity and other factors on correction outcomes when counter-attitudinal or neutral misinformation is encountered, where corrections may not directly contradict one's prior attitudes.

Also, although the manipulation of treating perceived credibility of familiar sources after exposure to correction as their prior source credibility perceptions was successful (see pages 83 & 84), it is likely that credibility perceptions of certain sources were influenced by the correction already, which may help explain the insignificant moderation effect of source credibility on the relationship between language intensity and message credibility of the correction. To address this limitation, future research can employ a two-phase study design to obtain individuals' true prior source credibility perceptions of the correction sources before being exposed to corrections with a time interval to avoid priming effects.

Additionally, the examination of corrections from familiar and unfamiliar sources could be extended by investigating the role of source familiarity in various relationships, along with the concept of credibility transfer, where the perceived credibility of a source is influenced by the quality of the message it delivers. This is particularly relevant in modern media environments where crowdsourced or algorithm-generated content plays an increasing role. Understanding how source familiarity affects source credibility, and how credibility transfers between source and message in these contexts could help improve the design of effective corrections.

In addition to the message feature (i.e., language intensity) and individual features (i.e., attitude discrepancy and issue involvement) examined in this study, future research should explore other important message factors, such as one-sided vs. two-sided arguments, positive vs. negative framing, and emotional tone. Investigating how these factors interact

with various source and audience factors could provide deeper insights into how different message delivery styles impact the effectiveness of corrections across diverse contexts. Other individual characteristics, such as cognitive ability, political ideology, and emotional susceptibility, should also be explored. Although age was not included as a covariate in this study given its insignificant correlations with main variables, it is highly likely that different age populations differ in their expectations or tolerance for intense language used in correction messages (Davenport et al., 2019). Future research could capture these differences by using a more age-diverse sample, as well as examining differences among other demographic groups such as education, income, and religion. Doing so would offer a more comprehensive understanding of how corrections are processed by different audience segments, especially in highly polarized or emotionally charged situations.

It is also worth noting that no backfire effects were observed in this study. While high-intensity language corrections were less effective than those using less intense language, they still reduced misinformation credibility compared to the control group across various feature groups. Backfire effects of correcting misinformation are of concern in the literature (Lewandowsky et al., 2012; Nyhan & Reifler, 2010; Wood & Porter, 2019) are especially relevant to correction wording and tone, as language that is perceived as too intense for a correction could potentially cause a backfire effect by undermining confidence in the truthfulness of the correction message or the credibility of the correction source. Future research should thus aim to replicate the findings in this dissertation and explore potential backfire effects under other conditions.

Finally, future research should employ more advanced methodologies, such as conjoint experiments, to examine the combined effects of multiple factors simultaneously.

This approach would allow researchers to manipulate and analyze various message, source, and individual characteristics simultaneously in a systematic way, enabling the identification of nuanced interactions that might not be captured in traditional experimental designs (Motta, 2022). It can also reveal the relative importance of different factors, helping to determine which features are most influential in shaping responses to corrections. Furthermore, the flexibility of conjoint experiments allows for the inclusion of a broader range of variables without trading off the statistical power (Knudsen & Johannesson, 2019), providing a richer and more comprehensive view of the factors affecting correction effectiveness.

Conclusion

In conclusion, this dissertation has shed light on the significant roles that features from all perspectives play in the effectiveness of misinformation corrections. By examining the interplay between language intensity, source credibility, attitude discrepancy and issue involvement, the research highlights the intricate dynamics involved in crafting effective corrections. The proposed language intensity and correction effectiveness model (LICE) illustrates that while corrections generally reduce belief in the misinformation, the use of high-intensity language may hinder effectiveness by undermining both message and source credibility, particularly when the source is unfamiliar. Even familiar, high-credibility sources are unable to mitigate this negative effect. This impact is especially pronounced among individuals with low attitude discrepancy or high issue involvement.

These findings emphasize the need for a more balanced approach in designing effective correction strategies. Tailoring corrections with appropriate language intensity to align with both the credibility of the source and the audience's pre-existing beliefs and involvement can enhance their effectiveness. The study offers practical implications for

improving debunking strategies for different contexts and audiences. The LICE model is important as a starting point for understanding the sets of factors that should be considered to better understand correction effectiveness. Future research should continue to explore how various factors interact across communication platforms and issue domains to optimize misinformation correction efforts.

References

Agadjanian, A., Bakhru, N., Chi, V., Greenberg, D., Hollander, B., Hurt, A., Kind, J., Lu, R., Ma, A., Nyhan, B., Pham, D., Qian, M., Tan, M., Wang, C., Wasdahl, A., & Woodruff, A. (2019). Counting the Pinocchios: The effect of summary fact-checking data on perceived accuracy and favorability of politicians. *Research & Politics*, *6*(3), 1-10.

Aird, M. J., Ecker, U. K., Swire, B., Berinsky, A. J., & Lewandowsky, S. (2018). Does truth matter to voters? The effects of correcting political misinformation in an Australian sample. *Royal Society Open Science*, *5*(12), 180593.

Albarracín, D., Kumkale, G. T., & Poyner-Del Vento, P. (2017). How people can become persuaded by weak messages presented by credible communicators: Not all sleeper effects are created equal. *Journal of Experimental Social Psychology*, *68*, 171-180.

Albarracín, D., & Vargas, P. (2010). Attitudes and persuasion: From biology to social responses to persuasive intent. In S. Fiske, D. Gilbert, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 394–427). Wiley.

Allen, J., Arechar, A. A., Pennycook, G., & Rand, D. G. (2021). Scaling up fact-checking using the wisdom of crowds. *Science Advances*, *7*(36), eabf4393.

Allen, M. (1991). Meta-analysis comparing the persuasiveness of one-sided and two-sided messages. *Western Journal of Speech Communication*, *55*(4), 390-404.

Amazeen, M. A. (2024). The misinformation recognition and response model: an emerging theoretical framework for investigating antecedents to and consequences of misinformation recognition. *Human Communication Research*, *50*(2), 218-229.

Amazeen, M. A., Thorson, E., Muddiman, A., & Graves, L. (2018). Correcting political and consumer misperceptions: The effectiveness and effects of rating scale versus contextual correction formats. *Journalism & Mass Communication Quarterly*, *95*(1), 28-48.

Anderson, N. H. (1959). Test of a model for opinion change. *The Journal of Abnormal and Social Psychology*, *59*(3), 371-381.

Appelman, A., & Sundar, S. S. (2016). Measuring message credibility: Construction and validation of an exclusive scale. *Journalism & Mass Communication Quarterly*, *93*(1), 59-79.

Autry, K. S., & Duarte, S. E. (2021). Correcting the unknown: Negated corrections may increase belief in misinformation. *Applied Cognitive Psychology*, *35*(4), 960-975.

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

Baek, T. H., Yoon, S., & Kim, S. (2015). When environmental messages should be assertive: Examining the moderating role of effort investment. *International Journal of Advertising*, *34*(1), 135-157.

Bak-Coleman, J. B., Kennedy, I., Wack, M., Beers, A., Schafer, J. S., Spiro, E. S., Starbird, K., & West, J. D. (2022). Combining interventions to reduce the spread of viral misinformation. *Nature Human Behaviour*, *6*(10), 1372-1380.

Banas, J. A., Palomares, N. A., Richards, A. S., Keating, D. M., Joyce, N., & Rains, S. A. (2022). When Machine and Bandwagon Heuristics Compete: Understanding Users'

Response to Conflicting AI and Crowdsourced Fact-Checking. *Human Communication Research*, *48*(3), 430-461.

Banerjee, B., & McKeage, K. (1994). How green is my value: exploring the relationship between environmentalism and materialism. In C. T. Allen & D. R. John (Eds.), *Advances in Consumer Research* (pp. 147-152). Association for Consumer Research.

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, *26*, 1531–1542.

Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W. P., & Linden, S. V. D. (2021). Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data & Society*, *8*(1), 20539517211013868.

Baum, J., & Abdel Rahman, R. (2021). Emotional news affects social judgments independent of perceived media credibility. *Social Cognitive and Affective Neuroscience*, *16*(3), 280-291.

Berinsky, A. J. (2017). Rumors and health care reform: Experiments in political misinformation. *British Journal of Political Science*, *47*(2), 241-262.

Bilandzic, H., & Busselle, R. (2013). Narrative persuasion. In J. P. Dillard, & L. Shen (Eds.), *The SAGE handbook of persuasion: Developments in theory and practice* (pp. 200-219). SAGE Publications.

Blankenship, K. L., Wegener, D. T., & Murray, R. A. (2015). Values, inter-attitudinal structure, and attitude change: Value accessibility can increase a related attitude's resistance to change. *Personality and Social Psychology Bulletin*, *41*(12), 1739-1750.

Bode, L., & Vraga, E. K. (2021). Correction experiences on social media during COVID-19. *Social Media+ Society*, *7*(2).

Bode, L., Vraga, E. K., & Tully, M. (2020). Do the right thing: Tone may not affect correction of misinformation on social media. *Harvard Kennedy School Misinformation Review*, *1*(4).

Bolsen, T., & Druckman, J. N. (2018). Do partisanship and politicization undermine the impact of a scientific consensus message about climate change?. *Group Processes & Intergroup Relations*, *21*(3), 389-402.

Borah, P., Kim, S., Xiao, X., & Lee, D. K. L. (2021). Correcting misinformation using theory-driven messages: HPV vaccine misperceptions, information seeking, and the moderating role of reflection. *Atlantic Journal of Communication*, *30*(3), 316-331.

Boukes, M., & Hameleers, M. (2022). Fighting lies with facts or humor: Comparing the effectiveness of satirical and regular fact-checks in response to misinformation and disinformation. *Communication Monographs*, 1-23.

Bowers, J. W. (1963). Language intensity, social introversion, and attitude change. *Communications Monographs*, *30*(4), 345-352.

Bradac, J. J., Bowers, J. W., & Courtright, J. A. (1980). Lexical variations in intensity, immediacy, and diversity: An axiomatic theory and causal model. In R. N. St. Clair & H. Giles (Eds.), *The social and psychological contexts of language* (pp. 193–223). Erlbaum.

Braddock, K., & Dillard, J. P. (2016). Meta-analytic evidence for the persuasive effect of narratives on beliefs, attitudes, intentions, and behaviors. *Communication Monographs*, *83*, 446–467.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, *114*(28), 7313-7318.

Brehm, J. W. (1966). *A theory of psychological reactance*. Academic Press.

Brydges, C. R., Gignac, G. E., & Ecker, U. K. (2018). Working memory capacity, short-term memory capacity, and the continued influence effect: A latent-variable analysis. *Intelligence*, *69*, 117-122.

Buchanan, T. (2020). Why do people spread false information online? The effects of message and viewer characteristics on self-reported likelihood of sharing social media disinformation. *Plos One*, *15*(10), e0239666.

Buller, D. B., Burgoon, M., Hall, J. R., Levine, N., Taylor, A. M., Beach, B. H., Melcher, C., Buller, M. K., Bowen, S. L., Hunsaker, F. G., & Bergen, A. (2000). Using language intensity to increase the success of a family intervention to protect children from ultraviolet radiation: predictions from language expectancy theory. *Preventive Medicine*, *30*(2), 103-113.

Burgers, C., & de Graaf, A. (2013). Language intensity as a sensationalistic news feature: The influence of style on sensationalism perceptions and effects. *Communications-The European Journal of Communication Research*, *38*(2), 167-188.

Burgoon, M., Denning, V. P., & Roberts, L. (2002). Language expectancy theory. In J. P. Dillard, & M. Pfau (Eds.), *The persuasion handbook: Developments in theory and practice* (pp. 117-136). Sage.

Burgoon, M., Jones, S. B., & Stewart, D. (1975). Toward a message-centered theory of persuasion: Three empirical investigations of language intensity. *Human Communication Research*, *1*(3), 240-256.

Burgoon, M., & Miller, G. R. (2018). An expectancy interpretation of language and persuasion. In H. Giles, & R. N. St. Clair (Eds.), *Recent advances in language, communication, and social psychology* (pp. 199-229). Routledge.

Cacioppo, J. T., & Petty, R. E. (1984). The elaboration likelihood model of persuasion. *ACR North American Advances*.

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.

Carnahan, D., & Bergan, D. E. (2022). Correcting the misinformed: The effectiveness of fact-checking messages in changing false beliefs. *Political Communication*, *39*(2), 166-183.

Carnahan, D., & Garrett, R. K. (2020). Processing style and responsiveness to corrective information. *International Journal of Public Opinion Research*, *32*(3), 530-546.

Carnahan, D., Hao, Q., Jiang, X., & Lee, H. (2018). Feeling fine about being wrong: The influence of self-affirmation on the effectiveness of corrective information. *Human Communication Research*, *44*(3), 274-298.

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.

Chan, M. P. 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.

Chen, X., Sin, S. C. J., Theng, Y. L., & Lee, C. S. (2015). Why students share misinformation on social media: Motivation, gender, and study-level differences. *The Journal of Academic Librarianship*, *41*(5), 583-592.

Chung, M., & Kim, N. (2021). When I learn the news is false: How fact-checking information stems the spread of fake news via third-person perception. *Human Communication Research*, *47*(1), 1-24.

Clark, A. R. (1993). The impact of cost of compliance, deservingness of aid, and directness of request on reactions to the request. *Southern Communication Journal*, *58*(3), 215–226.

Clark, T. (1998). The impact of candid versus legally defensible language on the persuasiveness of environmental self-assessments. *The Journal of Business Communication, 35*(3), 368-382.

Clementson, D. E., Pascual-Ferra, P., & Beatty, M. J. (2016). How language can influence political marketing strategy and a candidate's image: Effect of presidential candidates' language intensity and experience on college students' ratings of source credibility. *Journal of Political Marketing*, *15*(4), 388-415.

Cleveland, M., Kalamas, M., & Laroche, M. (2005). Shades of green: Linking environmental locus of control and pro-environmental behaviors. *Journal of Consumer Marketing*, *22*(4), 198-212.

Cook, J., & Lewandowsky, S. (2016). Rational irrationality: Modeling climate change belief polarization using Bayesian networks. *Topics in Cognitive Science*, *8*, 160–179.

Craig, T. Y., & Blankenship, K. L. (2011). Language and persuasion: Linguistic extremity influences message processing and behavioral intentions. *Journal of Language and Social Psychology*, *30*(3), 290-310.

Daft, R. L., & Lengel, R. H. (1984). Information richness: A new approach to managerial behavior and organizational design. *Research in Organizational Behavior*, *6*, 191–233.

Dai, S., Chen, K., & Jin, R. (2022). The effect of message framing and language intensity on green consumption behavior willingness. *Environment, Development and Sustainability*, *24*(2), 2432-2452.

Danielson, R. W., Sinatra, G. M., & Kendeou, P. (2016). Augmenting the refutation text effect with analogies and graphics. *Discourse Processes*, *53*(5-6), 392-414.

Davenport, S., Gizaw, M., Marcellino, W., Smith, S., Blake, J. S., & Kavanagh, J. (2019). *Facts versus opinions: How the style and language of news presentation is changing in the digital age*. RAND. https://www.rand.org/pubs/research_briefs/RB10059.html

De keersmaecker, J., & Roets, A. (2017). 'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. *Intelligence*, *65*, 107-110.

Denner, N., Viererbl, B., & Koch, T. (2023). Effects of repeated corrections of misinformation on organizational trust: more is not always better. *International Journal of Strategic Communication*, *17*(1), 39-53.

Dias, N., Pennycook, G., & Rand, D. G. (2020). Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media. *Harvard Kennedy School Misinformation Review*, *1*(1).

Dixon, G. N., McKeever, B. W., Holton, A. E., Clarke, C., & Eosco, G. (2015). The power of a picture: Overcoming scientific misinformation by communicating weight-of-evidence information with visual exemplars. *Journal of Communication*, *65*(4), 639-659.

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *Plos one*, *18*(3), e0279720.

Eagly, A. H. (1974). Comprehensibility of persuasive arguments as a determinant of opinion change. *Journal of Personality and Social Psychology*, *29*(6), 758-773.

Ecker, U. K., & Antonio, L. M. (2021). Can you believe it? An investigation into the impact of retraction source credibility on the continued influence effect. *Memory & Cognition*, *49*, 631-644.

Ecker, U. K., 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.

Ecker, U. K., Lewandowsky, S., & Chadwick, M. (2020). Can corrections spread misinformation to new audiences? Testing for the elusive familiarity backfire effect. *Cognitive Research: Principles and Implications*, *5*, 1-25.

Ecker, U. K., Lewandowsky, S., Cheung, C. S., & Maybery, M. T. (2015). He did it! She did it! No, she did not! Multiple causal explanations and the continued influence of misinformation. *Journal of Memory and Language*, *85*, 101-115.

Ecker, U. K., 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.

Ecker, U. K., Lewandowsky, S., Fenton, O., & Martin, K. (2014). Do people keep believing because they want to? Preexisting attitudes and the continued influence of misinformation. *Memory & Cognition*, *42*, 292-304.

Ecker, U. K., Lewandowsky, S., Jayawardana, K., & Mladenovic, A. (2019). Refutations of equivocal claims: No evidence for an ironic effect of counterargument number. *Journal of Applied Research in Memory and Cognition*, *8*(1), 98-107.

Ecker, U. K., Lewandowsky, S., Swire, B., & Chang, D. (2011). Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction. *Psychonomic Bulletin & Review*, *18*, 570-578.

Ecker, U. K., Lewandowsky, S., & Tang, D. T. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, *38*(8), 1087-1100.

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.

Epstein, Z., Foppiani, N., Hilgard, S., Sharma, S., Glassman, E., & Rand, D. (2022). Do explanations increase the effectiveness of AI-crowd generated fake news warnings?. In Y. Lin, M. Cha, & D. Quercia (Eds.), *Proceedings of the international AAAI conference on web and social media* (pp. 183-193). AAAI Press.

Eva, Q., Sakura, O., & Li, G. (2021). Mapping the field of misinformation correction and its effects: A review of four decades of research. *Social Science Information*, *60*(4), 522-547.

Fazio, L. (2020). Pausing to consider why a headline is true or false can help reduce the sharing of false news. *Harvard Kennedy School Misinformation Review*, *1*(2).

Feinberg, M., & Willer, R. (2013). The moral roots of environmental attitudes. *Psychological Science*, *24*(1), 56-62.

Ferracioli, P., Kniess, A. B., & Marques, F. P. J. (2022). The watchdog role of fact-checkers in different media systems. *Digital Journalism*, *10*(5), 717-737.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Flora, J. A., & Maibach, E. W. (1990). Cognitive responses to AIDS information: The effects of issue involvement and message appeal. *Communication Research*, *17*(6), 759-774.

Forgas, J. P., & East, R. (2008). On being happy and gullible: Mood effects on skepticism and the detection of deception. *Journal of Experimental Social Psychology*, *44*(5), 1362-1367.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25-42.

Fridkin, K., Kenney, P. J., & Wintersieck, A. (2015). Liar, liar, pants on fire: How fact-checking influences citizens' reactions to negative advertising. *Political Communication*, *32*(1), 127-151.

Garrett, R. K., Nisbet, E. C., & Lynch, E. K. (2013). Undermining the corrective effects of media-based political fact checking? The role of contextual cues and naïve theory. *Journal of Communication*, *63*(4), 617-637.

Garrett, R. K., & Weeks, B. E. (2013). The promise and peril of real-time corrections to political misperceptions. In A. S. Bruckman (Ed.), *Proceedings of the 2013 conference on computer supported cooperative work* (pp. 1047-1058). Association for Computing Machinery.

Gerend, M. A., & Shepherd, J. E. (2007). Using message framing to promote acceptance of the human papillomavirus vaccine. *Health Psychology*, *26*(6), 745–752.

Gerosa, T., Gui, M., Hargittai, E., & Nguyen, M. H. (2021). (Mis) informed during COVID-19: How education level and information sources contribute to knowledge gaps. *International Journal of Communication*, *15*, 2196-2217.

Graham, M. H. (2018). Self-awareness of political knowledge. *Political Behavior*, *42*, 305–326.

Graves, L., Nyhan, B., & Reifler, J. (2015). The diffusion of fact-checking. *American Press Institute*.

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.

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.

Gurgun, S., Cemiloglu, D., Close, E. A., Phalp, K., Nakov, P., & Ali, R. (2024). Why do we not stand up to misinformation? Factors influencing the likelihood of challenging misinformation on social media and the role of demographics. *Technology in Society*, *76*, 102444.

Haglin, K. (2017). The limitations of the backfire effect. *Research & Politics*, *4*(3), 1-5.

Hameleers, M. (2022). Separating truth from lies: Comparing the effects of news media literacy interventions and fact-checkers in response to political misinformation in the US and Netherlands. *Information, Communication & Society*, *25*(1), 110-126.

Hameleers, M., & Van der Meer, T. G. (2020). Misinformation and polarization in a high-choice media environment: How effective are political fact-checkers?. *Communication Research*, *47*(2), 227-250.

Hamilton, M. A., & Hunter, J. E. (1998). The effect of language intensity on receiver evaluations of message, source, and topic. In M. Allen & R. W. Preiss (Eds.), *Persuasion: Advances through meta- analysis* (pp. 99–138). Hampton Press.

Hamilton, M. A., Hunter, J. E., & Burgoon, M. (1990). An empirical test of an axiomatic model of the relationship between language intensity and persuasion. *Journal of Language and Social Psychology*, *9*(4), 235-255.

Hamilton, M. A., & Stewart, B. L. (1993). Extending an information processing model of language intensity effects. *Communication Quarterly*, *41*(2), 231-246.

Hart, P. S., & Nisbet, E. C. (2012). Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies. *Communication Research*, *39*(6), 701-723.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.

Heiss, R., Nanz, A., Knupfer, H., Engel, E., & Matthes, J. (2023). Peer correction of misinformation on social media:(In) civility, success experience and relationship consequences. *New Media & Society*, 14614448231209946.

Holman, M. R., & Lay, J. C. (2019). They see dead people (voting): Correcting misperceptions about voter fraud in the 2016 US presidential election. *Journal of Political Marketing*, *18*(1-2), 31-68.

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion*. Yale University Press.

Hovland, C. I., Lumsdaine, A. A., & Sheffield, F. D. (1949). *Experiments on mass communication*. Princeton University Press.

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly, 15*(4), 635-650.

Huang, Y., & Wang, W. (2022). When a story contradicts: correcting health misinformation on social media through different message formats and mechanisms. *Information, Communication & Society*, *25*(8), 1192-1209.

Huber, B., Borah, P., & Gil de Zúñiga, H. (2022). Taking corrective action when exposed to fake news: The role of fake news literacy. *Journal of Media Literacy Education*, *14*(2), 1-14.

Hunter, J. E., Danes, J. E., & Cohen, S. H. (1984). *Mathematical models of attitude change: Change in single attitudes and cognitive structure* (Vol. 1). Academic Press.

ISideWith. (2023). *Political issues of 2023*. https://www.isidewith.com/polls

Iswara, A. A., & Bisena, K. A. (2020). Manipulation And persuasion through language features in fake news. *RETORIKA: Jurnal Ilmu Bahasa*, *6*(1), 26-32.

Jamieson, K. H., & Hardy, B. W. (2014). Leveraging scientific credibility about Arctic sea ice trends in a polarized political environment. *Proceedings of the National Academy of Sciences, 111*(Suppl. 4), 13598–13605.

Jenke, L. (2023). Affective polarization and misinformation belief. *Political Behavior*, 1-60.

Jin, H., Jia, L., Yin, X., Wei, S., & Xu, G. (2022). The influence of information relevance on the continued influence effect of misinformation. *Journal of Psychiatry and Psychiatric Disorders*, *6*(3), 203-218.

Jin, Y., van der Meer, T. G., Lee, Y. I., & Lu, X. (2020). The effects of corrective communication and employee backup on the effectiveness of fighting crisis misinformation. *Public Relations Review*, *46*(3), 101910.

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.

Jones-Jang, S. M., Mortensen, T., & Liu, J. (2021). Does media literacy help identification of fake news? Information literacy helps, but other literacies don't. *American Behavioral Scientist*, *65*(2), 371-388.

Kan, I. P., Pizzonia, K. L., Drummey, A. B., & Mikkelsen, E. J. (2021). Exploring factors that mitigate the continued influence of misinformation. *Cognitive Research: Principles and Implications*, *6*(1), 1-33.

Kim, Y., Baek, T. H., Yoon, S., Oh, S., & Choi, Y. K. (2017). Assertive environmental advertising and reactance: Differences between South Koreans and Americans. *Journal of Advertising*, *46*(4), 550-564.

Kligler-Vilenchik, N. (2022). Collective social correction: Addressing misinformation through group practices of information verification on WhatsApp. *Digital Journalism*, *10*(2), 300-318.

Knudsen, E., & Johannesson, M. P. (2019). Beyond the limits of survey experiments: How conjoint designs advance causal inference in political communication research. *Political Communication*, *36*(2), 259-271.

Kronrod, A., Grinstein, A., & Wathieu, L. (2012). Go green! Should environmental messages be so assertive?. *Journal of Marketing*, *76*(1), 95-102.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480-498.

Lang, A., Dhillon, K., & Dong, Q. (1995). The effects of emotional arousal and valence on television viewers' cognitive capacity and memory. *Journal of Broadcasting & Electronic Media*, *39*, 313–327.

Lee, J. (2022). The effect of web add-on correction and narrative correction on belief in misinformation depending on motivations for using social media. *Behaviour & Information Technology*, *41*(3), 629-643.

Lee, J., Kim, J. W., & Yun Lee, H. (2022). Unlocking conspiracy belief systems: How fact-checking label on Twitter counters conspiratorial MMR vaccine misinformation. *Health Communication*, 38(9), 1780-1792.

Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E., Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., … & Zaragoza, M. S. (2020). The Debunking Handbook 2020. Doi: 10.17910/b7.1182

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106-131.

Li, J., & Wagner, M. W. (2020). The value of not knowing: Partisan cue-taking and belief updating of the uninformed, the ambiguous, and the misinformed. *Journal of Communication*, *70*(5), 646-669.

Li, M. H., Chen, Z., & Rao, L. L. (2022). Emotion, analytic thinking and susceptibility to misinformation during the COVID-19 outbreak. *Computers in Human Behavior*, *133*, 107295.

Liebrecht, C., Hustinx, L., & Van Mulken, M. (2019). The relative power of negativity: the influence of language intensity on perceived strength. *Journal of Language and Social Psychology*, *38*(2), 170-193.

Lim, S., & Kwon, N. (2010). Gender differences in information behavior concerning Wikipedia, an unorthodox information source?. *Library & Information Science Research*, *32*(3), 212-220.

Liu, X., Qi, L., Wang, L., & Metzger, M. J. (2023). Checking the fact-checkers: The role of source type, perceived credibility, and individual differences in fact-checking effectiveness. *Communication Research*, 00936502231206419.

Lyons, B., Mérola, V., Reifler, J., & Stoeckel, F. (2020). How politics shape views toward fact-checking: Evidence from six European countries. *The International Journal of Press/Politics*, *25*(3), 469-492.

Ma, H., & Miller, C. (2022). "I felt completely turned off by the message": The effects of controlling language, fear, and disgust appeals on responses to COVID-19 vaccination messages. *Journal of Health Communication*, *27*(6), 427-438.

Marshall, H. M., Reinhart, A. M., Feeley, T. H., Tutzauer, F., & Anker, A. (2008). Comparing college students' value-, outcome-, and impression-relevant involvement in health-related issues. *Health Communication*, *23*(2), 171-183.

Martel, C., Mosleh, M., & Rand, D. G. (2021). You're definitely wrong, maybe: Correction style has minimal effect on corrections of misinformation online. *Media and Communication*, *9*(1), 120-133.

Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications*, *5*(47), 1-20.

McEwen, W. J., & Greenberg, B. S. (1970). The effects of message intensity on receiver evaluations of source, message and topic. *Journal of Communication*, *20*(4), 340-350.

McIlhiney, P., Gignac, G. E., Weinborn, M., & Ecker, U. K. (2022). Sensitivity to misinformation retractions in the continued influence paradigm: Evidence for stability. *Quarterly Journal of Experimental Psychology*, *75*(7), 1259-1271.

Metzger, M. J., Qi, L., Liu, X., & Wang, L. (2023, January). *Facts about fact-checkers: Comparing public perceptions and effectiveness across fact-checking sources*. [Paper presentation]. The Misinformation and its Consequences for American Democracy Conference at the Jack Citrin Center for Public Opinion Research, UC Berkeley.

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication*, *60*(3), 413-439.

Miller, C. H., Lane, L. T., Deatrick, L. M., Young, A. M., & Potts, K. A. (2007). Psychological reactance and promotional health messages: The effects of controlling language, lexical concreteness, and the restoration of freedom. *Human Communication Research*, *33*(2), 219-240.

Miller, J. M., Saunders, K. L., & Farhart, C. E. (2016). Conspiracy endorsement as motivated reasoning: The moderating roles of political knowledge and trust. *American Journal of Political Science*, *60*(4), 824–844.

Motta, M. (2022). Advancing the study of misinformation correction through conjoint experimentation. https://osf.io/8hnyg/download

Murphy, S. T., Frank, L. B., Chatterjee, J. S., Moran, M. B., Zhao, N., Amezola de Herrera, P., & Baezconde-Garbanati, L. A. (2015). Comparing the relative efficacy of narrative vs nonnarrative health messages in reducing health disparities using a randomized trial. *American Journal of Public Health*, *105*, 2117–2123.

Nabi, R. L., Moyer-Gusé, E., & Byrne, S. (2007). All joking aside: A serious investigation into the persuasive effect of funny social issue messages. *Communication Monographs*, *74*(1), 29-54.

Nan, X. (2012). Communicating to young adults about HPV vaccination: Consideration of message framing, motivation, and gender. *Health Communication*, *27*(1), 10–18.

Nan, X., & Madden, K. (2012). HPV vaccine information in the blogosphere: how positive and negative blogs influence vaccine-related risk perceptions, attitudes, and behavioral intentions. *Health Communication*, *27*(8), 829-836.

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303-330.

Nyhan, B., & Reifler, J. (2015). Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. *Vaccine*, *33*(3), 459-464.

Nyhan, B., Reifler, J., Richey, S., & Freed, G. L. (2014). Effective messages in vaccine promotion: a randomized trial. *Pediatrics*, *133*(4), e835-e842.

Nyhan, B., Reifler, J., & Ubel, P. A. (2013). The hazards of correcting myths about health care reform. *Medical Care*, *51*, 127–132.

Oeldorf-Hirsch, A., Schmierbach, M., Appelman, A., & Boyle, M. P. (2020). The ineffectiveness of fact-checking labels on news memes and articles. *Mass Communication and Society*, *23*(5), 682-704.

O'Keefe, D. J. (1990). *Persuasion: Theory and research*. Sage Publications.

O'Keefe, D. J. (1999). How to handle opposing arguments in persuasive messages: A meta-analytic review of the effects of one-sided and two-sided messages. *Annals of the International Communication Association*, *22*(1), 209-249.

Okuno, H., Arai, S., Suzuki, M., & Kikkawa, T. (2022). Impact of Refutational Two-Sided Messages on Attitudes Toward Novel Vaccines Against Emerging Infectious Diseases During the COVID-19 Pandemic. *Frontiers in Public Health*, *10*, 775486.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin Books UK.

Pasek, J., Sood, G., & Krosnick, J. A. (2015). Misinformed about the affordable care act? Leveraging certainty to assess the prevalence of misperceptions: Misinformed about the affordable care act. *Journal of Communication*, *65*(4), 660–673.

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). Everyday consequences of analytic thinking. *Current Directions in Psychological Science*, *24*(6), 425-432.

Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences*, *116*(7), 2521-2526.

Pennycook, G., & Rand, D. G. (2022). Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation**.** *Nature Communications, 13*(2333).

Petty, R. E., & Cacioppo, J. T. (1984). The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology*, *46*, 69–81.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In R. E. Petty, & J. T. Cacioppo (Eds.), *Communication and persuasion: Central and peripheral routes to attitude change* (pp. 1-24). Springer New York.

Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology*, *41*(5), 847.

Petty, R. E., Cacioppo, J. T., & Schumann, D. (1983). Central and peripheral routes to advertising effectiveness: The moderating role of involvement. *Journal of Consumer Research*, *10*(2), 135-146.

Pew Research Center. (2024, September 17). *Social media and news fact sheet*. https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/

Pluviano, S., Della Sala, S., & Watt, C. (2020). The effects of source expertise and trustworthiness on recollection: the case of vaccine misinformation. *Cognitive Processing*, *21*(3), 321-330.

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, *34*, 243–281.

Porter, E., Velez, Y., & Wood, T. J. (2022). Factual corrections eliminate false beliefs about COVID-19 vaccines. *Public Opinion Quarterly*, *86*(3), 762-773.

Quick, B. L., & Considine, J. R. (2008). Examining the use of forceful language when designing exercise persuasive messages for adults: A test of conceptualizing reactance arousal as a two-step process. *Health Communication*, *23*(5), 483-491.

Quick, B. L., & Stephenson, M. T. (2007). Further evidence that psychological reactance can be modeled as a combination of anger and negative cognitions. *Communication Research*, *34*(3), 255-276.

Renkema, J. (1997). Geïntensiveerd taalgebruik: Een analyseschema [Intensified language: An analysis scheme]. In H. Van den Bergh, D. Janssen, N. Bertens, & M. Damen (Eds.), *Taalgebruik ontrafeld* [Language use explained] (pp. 495-504). Foris.

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.

Rieh, S. Y., & Danielson, D. R. (2007). Credibility: A multidisciplinary framework. In B. Cronin (Ed.), *Annual review of information science and technology* (Vol. 41, pp. 307–364). Medford, NJ: Information Today.

Robertson, C. T., Mourão, R. R., & Thorson, E. (2020). Who uses fact-checking sites? The impact of demographics, political antecedents, and media use on fact-checking site awareness, attitudes, and behavior. *The International Journal of Press/Politics*, *25*(2), 217-237.

Roozenbeek, J., Culloty, E., & Suiter, J. (2023). Countering misinformation: evidence, knowledge gaps, and implications of current interventions. *European Psychologist*, *28*(3), 189–205.

Roozenbeek, J., Freeman, A. L., & Van Der Linden, S. (2021). How accurate are accuracy-nudge interventions? A preregistered direct replication of Pennycook et al. (2020). *Psychological Science*, *32*(7), 1169-1178.

Roozenbeek, J., & van der Linden, S. (2020). Breaking Harmony Square: A game that "inoculates" against political misinformation. *Harvard Kennedy School Misinformation Review*, *1*(8).

Saleh, N., Roozenbeek, J., Makki, F., McClanahan, W., & van der Linden, S. (2021). Active inoculation boosts attitudinal resistance against extremist persuasion techniques – A novel approach towards the prevention of violent extremism. *Behavioural Public Policy*, 1–24.

Sanderson, J. A., Bowden, V., Swire-Thompson, B., Lewandowsky, S., & Ecker, U. K. (2022). 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.

Sanderson, J. A., & Ecker, U. K. (2020). The challenge of misinformation and ways to reduce its impact. In P. Van Meteris, A. List, D. Lombardi, & P. Kendeou (Eds.), *Handbook of learning from multiple representations and perspectives* (pp. 461-476). Taylor & Francis.

Sangalang, A., Ophir, Y., & Cappella, J. N. (2019). The potential for narrative correctives to combat misinformation. *Journal of Communication*, *69*(3), 298-319.

Schmid, P., & Betsch, C. (2019). Effective strategies for rebutting science denialism in public discussions. *Nature Human Behaviour*, *3*(9), 931-939.

Schultz, F., Utz, S., & Göritz, A. (2011). Is the medium the message? Perceptions of and reactions to crisis communication via Twitter, blogs and traditional media. *Public Relations Review*, *37*, 20–27.

Schwarz, N., Newman, E., & Leach, W. (2016). Making the truth stick & the myths fade: Lessons from cognitive psychology. *Behavioral Science & Policy*, *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. *Advances in Experimental Social Psychology*, *39*, 127-161.

Schweiger, W. (2000). Media credibility – experience or image?: A survey on the credibility of the world wide web in Germany in comparison to other media. *European Journal of Communication*, *15*, 37-59.

Sherman, D. K., & Cohen, G. L. (2002). Accepting threatening information: Self–Affirmation and the reduction of defensive biases. *Current Directions in Psychological Science*, *11*(4), 119-123.

Shin, J., & Thorson, K. (2017). Partisan selective sharing: The biased diffusion of fact-checking messages on social media. *Journal of Communication*, *67*(2), 233-255.

Shu, S. B., & Carlson, K. A. (2014). When three charms but four alarms: Identifying the optimal number of claims in persuasion settings. *Journal of Marketing*, *78*, 127–139.

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.

Slater, M. D., & Rouner, D. (1996). How message evaluation and source attributes may influence credibility assessment and belief change. *Journalism & Mass Communication Quarterly*, *73*(4), 974-991.

Sui, Y. J., & Zhang, B. (2021). Determinants of the perceived credibility of rebuttals concerning health misinformation. *International Journal of Environmental Research and Public Health, 18*(3), 1345.

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. Metzger & A. Flanagin (Eds.), *Digital Media, Youth, and Credibility* (pp. 73-100). The MIT Press.

Swire, B., Berinsky, A. J., Lewandowsky, S., & Ecker, U. K. (2017). Processing political misinformation: Comprehending the Trump phenomenon. *Royal Society Open Science*, *4*(3), 1-21.

Swire, B., Ecker, U. K., & Lewandowsky, S. (2017). The role of familiarity in correcting inaccurate information. Journal of experimental psychology: learning, memory, and cognition, 43(12), 1948-1961.

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.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*(3), 755-769.

Tandoc, E. C. Jr, Lim, Z. W., & Ling, R. (2018). Defining "fake news" A typology of scholarly definitions. *Digital Journalism*, *6*(2), 137-153.

Tannenbaum, M. B., Hepler, J., Zimmerman, R. S., Saul, L., Jacobs, S., Wilson, K., & Albarracín, D. (2015). Appealing to fear: A meta-analysis of fear appeal effectiveness and theories. *Psychological Bulletin*, *141*(6), 1178-1204.

Tay, L. Q., Hurlstone, M. J., Kurz, T., & Ecker, U. K. (2022). A comparison of prebunking and debunking interventions for implied versus explicit misinformation. *British Journal of Psychology*, *113*(3), 591-607.

Thorson, E., & Li, J. (2021). The effect of corrections and corrected misinformation. In H. Tumber, & S. R. Waisbord (Eds.), *The Routledge companion to media disinformation and populism* (pp. 548-558). Taylor & Francis.

Tormala, Z. L., Briñol, P., & Petty, R. E. (2006). When credibility attacks: The reverse impact of source credibility on persuasion. *Journal of experimental social psychology*, *42*(5), 684-691.

Traberg, C. S., & van der Linden, S. (2022). Birds of a feather are persuaded together: Perceived source credibility mediates the effect of political bias on misinformation susceptibility. *Personality and Individual Differences*, *185*, 111269.

Tseng, Y. C., Bi, N., Chang, Y. J., & Yuan, C. W. (2022). Investigating Correction Effects of Different Modalities for Misinformation about COVID-19. In G. Hsieh, A. Tang, M. G. Ames, S. Ding, S. Fussell, V. Liao, A. Monroy-Hernández, S. Munson, I. Shklovski, & J. Tang (Eds.), *Companion publication of the 2022 conference on computer supported cooperative work and social computing* (pp. 54-58). Association for Computing Machinery.

Vafeiadis, M., & Xiao, A. (2021). Fake news: How emotions, involvement, need for cognition and rebuttal evidence (story vs. informational) influence consumer reactions toward a targeted organization. *Public Relations Review*, *47*(4), 102088.

Van Bavel, J. J., Harris, E. A., Pärnamets, P., Rathje, S., Doell, K. C., & Tucker, J. A. (2021). Political psychology in the digital (mis) information age: A model of news belief and sharing. *Social Issues and Policy Review*, *15*(1), 84-113.

Van der Linden, S. (2022). Misinformation: susceptibility, spread, and interventions to immunize the public. *Nature medicine*, *28*(3), 460-467.

Van der Linden, S. L., Leiserowitz, A. A., Feinberg, G. D., & Maibach, E. W. (2014). How to communicate the scientific consensus on climate change: plain facts, pie charts or metaphors?. *Climatic Change*, *126*, 255-262.

Van der Meer, T. G., & Jin, Y. (2020). Seeking formula for misinformation treatment in public health crises: The effects of corrective information type and source. *Health Communication*, *35*(5), 560-575.

Vegetti, F., & Mancosu, M. (2020). The impact of political sophistication and motivated reasoning on misinformation. *Political Communication*, *37*(5), 678-695.

Vigderman, A. (2024, September 26). *90% of people claim they fact-check news stories as trust in media plummets*. Security.org. https://www.security.org/digital-security/misinformation-disinformation-survey/

Vitriol, J. A., Sandor, J., Vidigal, R., & Farhart, C. (2022). On the Independent Roles of Cognitive & Political Sophistication: Variation Across Attitudinal Objects. *Applied Cognitive Psychology*, *37*(2), 319-331.

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. (2017a). Using expert sources to correct health misinformation in social media. *Science Communication*, *39*(5), 621-645.

Vraga, E. K., & Bode, L. (2017b). I do not believe you: How providing a source corrects health misperceptions across social media platforms. *Information, Communication & Society*, *21*(10), 1337-1353.

Vraga, E. K., Bode, L., & Tully, M. (2022). Creating news literacy messages to enhance expert corrections of misinformation on Twitter. *Communication Research*, *49*(2), 245-267.

Vraga, E. K., Kim, S. C., & Cook, J. (2019). Testing logic-based and humor-based corrections for science, health, and political misinformation on social media. *Journal of Broadcasting & Electronic Media*, *63*(3), 393-414.

Vraga, E. K., Kim, S. C., Cook, J., & Bode, L. (2020). Testing the effectiveness of correction placement and type on Instagram. *The International Journal of Press/Politics*, *25*(4), 632-652.

Wagner, B. C., & Petty, R. E. (2011). The elaboration likelihood model of persuasion: Thoughtful and non-thoughtful social influence. In D. Chadee (Ed.), *Theories in social psychology* (pp. 96-116). Oxford.

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.

Walter, N., Brooks, J. J., Saucier, C. J., & Suresh, S. (2020). Evaluating the impact of attempts to correct health misinformation on social media: A meta-analysis. *Health Communication*, *36*(13), 1776-1784.

Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2019). Fact-checking: A meta-analysis of what works and for whom. *Political Communication*, *37*(3), 350-375.

Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, *85*(3), 423-441.

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.

Wang, W., & Huang, Y. (2021). Countering the "harmless e-cigarette" myth: The interplay of message format, message sidedness, and prior experience with e-cigarette use in misinformation correction. *Science Communication*, *43*(2), 170-198.

Wang, X., Chao, F., Yu, G., & Zhang, K. (2022). Factors influencing fake news rebuttal acceptance during the COVID-19 pandemic and the moderating effect of cognitive ability. *Computers in Human Behavior*, *130*, 107174.

Wang, Z., Walther, J. B., Pingree, S., & Hawkins, R. P. (2008). Health information, credibility, homophily, and influence via the Internet: Web sites versus discussion groups. *Health Communication*, *23*(4), 358–368.

Weeks, B. E. (2015). Emotions, partisanship, and misperceptions: How anger and anxiety moderate the effect of partisan bias on susceptibility to political misinformation. *Journal of Communication*, *65*(4), 699-719.

Wieluch, M. I. (2017). When risk communication messages should be assertive: Differentiating the moderating effects of perceived issue importance and perceived susceptibility. In V. Zabkar, & M. Eisend (Eds.), *Advances in Advertising Research VIII: Challenges in an Age of Dis-Engagement* (pp. 283-293). Springer Fachmedien Wiesbaden.

Wilkes, A. L., & Leatherbarrow, M. (1988). Editing episodic memory following the identification of error. *The Quarterly Journal of Experimental Psychology*, *40*(2), 361-387.

Winter, S., Krämer, N. C., Rösner, L., & Neubaum, G. (2015). Don't keep it (too) simple: How textual representations of scientific uncertainty affect laypersons' attitudes. *Journal of Language and Social Psychology*, *34*(3), 251-272.

Wogalter, M. S. (2006). Communication-human information processing (C-HIP) Model. In M. S. Wogalter (Ed.), *Handbook of warnings* (pp. 51–61). Lawrence Erlbaum.

Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, *41*(1), 135-163.

Xiao, X., & Su, Y. (2021). Integrating reasoned action approach and message sidedness in the era of misinformation: the case of HPV vaccination promotion. *Journal of Health Communication*, *26*(6), 371-380.

Xiao, X., Su, Y., & Lee, D. K. L. (2021). Who consumes new media content more wisely? Examining personality factors, SNS use, and new media literacy in the era of misinformation. *Social Media+ Society*, *7*(1), 2056305121990635.

Xue, H. (2021). *How do language intensity and artificial intelligence (AI) affect perceptions of fact-checking messages and evaluations of fact-checking agencies?* [Master's Thesis, UC Davis]. UC Davis Electronic Theses and Dissertations.

Yaqub, W., Kakhidze, O., Brockman, M. L., Memon, N., & Patil, S. (2020). Effects of credibiliity indicators on social media news sharing intent. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-14). Association for Computing Machinery.

Young, D. G., Jamieson, K. H., Poulsen, S., & Goldring, A. (2018). Fact-checking effectiveness as a function of format and tone: Evaluating FactCheck.org and FlackCheck.org. *Journalism & Mass Communication Quarterly*, *95*(1), 49–75.

Yu, W., Payton, B., Sun, M., Jia, W., & Huang, G. (2023). Toward an integrated framework for misinformation and correction sharing: A systematic review across domains. *New Media & Society*, *25*(8), 2241-2267.

Zhang, J., Featherstone, J. D., Calabrese, C., & Wojcieszak, M. (2021). Effects of fact-checking social media vaccine misinformation on attitudes toward vaccines. *Preventive Medicine*, *145*, 106408.

Zhang, Y., Guo, B., Ding, Y., Liu, J., Qiu, C., Liu, S., & Yu, Z. (2022). Investigation of the determinants for misinformation correction effectiveness on social media during COVID-19 pandemic. *Information Processing & Management*, *59*(3), 102935.

Zhou, L., Twitchell, D. P., Qin, T., Burgoon, J. K., & Nunamaker, J. F. (2003, January). An exploratory study into deception detection in text-based computer-mediated communication. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*. IEEE.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, *6*, 292–297.

Appendix A Textual Stimuli

Immigration

Anti-immigration news story:

Undocumented Immigrants Strain U.S. Welfare System, Costing $113 Billion Annually
According to a recent report by the Center for Immigration Studies (CIS), undocumented immigrants are accessing U.S. public welfare programs, incurring a significant financial burden on American taxpayers.

It has been estimated that these costs amount to a substantial $113 billion annually. This financial outlay is mostly from the utilization of welfare benefits, including but not limited to food stamps, Medicaid, public housing assistance, and cash assistance programs.

Data from the Coalition for Immigration Reform further confirms the prevalence of undocumented immigrants accessing food stamps and public housing assistance, emphasizing the challenge posed by unauthorized immigrants' reliance on public welfare programs.

Click the link in bio for more.

Fact-Check: The True Impact of Undocumented Immigrants on U.S. Public Welfare and the Economy

(low intensity)
A recent story posted by USA Today claims that undocumented immigrants strain the U.S. welfare system, costing $113 billion annually.

This story contains some inaccuracies. The CIS report did not indicate a significant financial burden from undocumented immigrants accessing U.S. public welfare programs. Instead, it described some general concerns by the public about this issue that are not supported by data.

The U.S. Immigration and Nationality Act (INA) prevents undocumented immigrants from accessing many benefits, including welfare, food assistance, non-emergency health coverage, disability coverage, and public or assisted housing.

At the same time, some economists have suggested that undocumented immigrants make contributions to the U.S. economy through their labor force participation, tax payments, and entrepreneurial endeavors.

Data from the U.S. Census Bureau indicates that undocumented immigrants contribute billions of dollars annually in state and local taxes, which may outweigh the costs associated with the public services they can access.

Read more at the link in bio.

(high intensity)
A recent story posted by USA Today claims that undocumented immigrants strain the U.S. welfare system, costing $113 billion annually.

This story is FALSE! The CIS report definitely did NOT indicate a significant financial burden from undocumented immigrants accessing U.S. public welfare programs. Instead, it only described some general concerns by the public about this issue that are not supported by ANY data.

IN FACT, the U.S. Immigration and Nationality Act (INA) explicitly BANS undocumented immigrants from accessing nearly all benefits, including welfare, food assistance, non-emergency health coverage, disability coverage, and public or assisted housing.

At the same time, economists have proved that undocumented immigrants make substantial contributions to the U.S. economy through their labor force participation, tax payments, and entrepreneurial endeavors.

Data from the U.S. Census Bureau strongly indicates that undocumented immigrants contribute billions of dollars annually in state and local taxes, which greatly outweighs the costs associated with the very few public services they can access. FAKE NEWS like this story should be banned!

Read more at the link in bio.

Pro-immigration news story:

Increased Immigration Boosts Economic Growth and Entrepreneurship
According to a recent report by the Center for Immigration Studies (CIS), increased immigration has boosted the country's GDP by a staggering 10% in the past decade.

Contrary to popular belief, immigrants have emerged as significant contributors to entrepreneurship and innovation in the U.S. The report suggests that immigrants are twice as likely to start businesses compared to native-born citizens, which has fueled economic recovery and job creation across various sectors.

The report also emphasizes the transformative role of immigrants in driving technological innovation, particularly in the AI and renewable energy sectors. Statistics from the National Foundation for American Policy (NFAP) confirms this assertion, revealing a remarkable 15% surge in technological advancements attributable to immigrant contributions.

Click the link in bio for more.

Fact-Check: The True Impact of Immigration on GDP and Technological Advancements

(low intensity)
A recent story posted by USA Today claims that increased immigration has boosted the country's GDP and technological advancements.

This story contains some inaccuracies. The CIS report did not provide specific statistics of an increase in GDP and technological advancements by immigration. Instead, it suggested potential contributions of immigrants to the economy.

Attributing 10% of GDP growth solely to immigration likely oversimplifies the complex factors that influence economic growth. Data from the U.S. Census Bureau suggests that immigration's contribution to GDP in the past decade is estimated to be 1.2-1.6%.

While immigrants may exhibit higher rates of entrepreneurship in certain contexts, this tendency is not universally consistent across regions and industries, which is influenced by factors including economic conditions, access to resources, and cultural background.

The claimed unique contribution of immigrants in technological sectors may also be exaggerated, as other factors should be considered too, such as domestic policies and investment trends.

Read more at the link in bio.

(high intensity)
A recent story posted by USA Today claims that increased immigration has boosted the country's GDP and technological advancements.

This story is FALSE! The CIS report definitely did NOT provide specific statistics of an increase in GDP and technological advancements by immigration. Instead, it merely suggested potential contributions of immigrants to the economy.

Attributing precisely 10% of GDP growth exclusively to immigration greatly oversimplifies the highly complex factors that influence economic growth. IN FACT, data from the U.S. Census Bureau strongly suggests that immigration's contribution to GDP in the past decade is only estimated to be 1.2-1.6%.

While immigrants exhibit higher rates of entrepreneurship in certain contexts, this tendency is NOT universally consistent across all regions and industries, which is influenced by many factors including economic conditions, access to resources, and cultural background.

The claimed unique contribution of immigrants in technological sectors is also completely unfounded, as many other factors MUST be considered too, such as domestic policies and investment trends. FAKE NEWS like this story should be banned!

Read more at the link in bio.

Abortion

Anti-abortion news story:

Alarming Health Risks Tied to Abortions: Increased Infertility and Mental Health Issues
A recent longitudinal study cited by the National Institutes of Health (NIH) reveals alarming statistics linking abortions to severe health complications, suggesting a 15% increase in infertility rates among women who have undergone the procedure.

The report also suggests a correlation between abortions and heightened mental health issues, revealing a staggering 20% rise in anxiety and depression cases among individuals who have had abortions.

Studies by advocacy groups have reported similar findings, raising important questions about long-term consequences of abortion that need further investigation.

Click the link in bio for more.

Fact-Check: The True Impact of Abortions on Infertility and Mental Health

(low intensity)
A recent story posted by USA Today claims that abortions are linked to increased infertility rates and mental health issues.

This story contains some inaccuracies. The NIH study cited did not report statistics linking abortions to a 15% increase in infertility rates. Instead, it suggested a potential rise in infertility rates due to incomplete abortions or procedures conducted by unqualified individuals. No scientific consensus supports a link between abortion and infertility when it is performed by qualified healthcare professionals.

Moreover, the relationship between mental health and infertility is complex. While the study indicates that women with poor mental health may have lower fertility, little evidence shows abortion is related to a 20% rise in cases of anxiety and depression.

Studies published in medical journals have also recognized other factors that should also be considered. These include pre-existing mental health risks, socioeconomic status, and access to public welfare, all of which may influence mental health outcomes in addition to abortion.

Read more at the link in bio.

(high intensity)
A recent story posted by USA Today claims that abortions are linked to increased infertility rates and mental health issues.

This story is FALSE! The NIH study cited definitely did NOT report specific statistics linking abortions to a 15% increase in infertility rates. Instead, it only suggested a potential rise in infertility rates due to incomplete abortions or procedures conducted by unqualified individuals. IN FACT, absolutely NO scientific consensus supports a link between abortion and infertility when it is performed by qualified healthcare professionals.

Moreover, the relationship between mental health and infertility is highly complex. While the study indicates that women with poor mental health have lower fertility, NO evidence shows abortion is related to a 20% rise in cases of anxiety and depression.

Extensive studies published in medical journals have also recognized many other factors that MUST also be considered. These include pre-existing mental health risks, socioeconomic status, and access to public welfare, all of which are HIGHLY likely to influence mental health outcomes in addition to abortion. FAKE NEWS like this story should be banned!

Read more at the link in bio.

Pro-abortion news story:

Increased Abortion Access Linked to Decline in Poverty and Crime Rates
A recent longitudinal study cited by the National Institutes of Health (NIH) highlights positive outcomes correlated with increased access to abortion services.

According to the findings, regions where abortion services have been made more accessible have experienced a remarkable 14% decrease in poverty rates. There is also a notable 18% decline in crime rates in areas where abortion availability has grown, implying the positive societal impacts of abortion access.

Data from the Guttmacher Institute, a leading research organization on reproductive health, further confirms that individuals who are denied access to abortion experience significantly more adverse socio-economic outcomes, including increased poverty rates and limited educational and employment opportunities.

Click the link in bio for more.

Fact-Check: The True Impact of Abortion Access on Poverty and Crime Rates

(low intensity)
A recent story posted by USA Today claims that increased abortion access is linked to a decline in poverty and crime rates.

This story contains some inaccuracies. The NIH study cited did not report specific statistics linking increased abortion access to a 14% decrease in poverty rates. Instead, it suggested a potential decline in poverty rates as a result of various socioeconomic factors, including access to reproductive healthcare services.

Data from the U.S. Census Bureau suggests that access to abortion may be associated with a decline in the number of children living in poverty by about 1.2% over the past 20 years.

Little evidence shows increased abortion access is related to a 18% decline in crime rates. Other factors that may influence crime rates over time should be considered, such as law enforcement policies and community interventions.

The findings from the Guttmacher Institute are also subject to interpretation and may not universally apply to all contexts, which could vary depending on factors including geographic location, demographic characteristics, and methodological approaches.

Read more at the link in bio.

(high intensity)
A recent story posted by USA Today claims that increased abortion access is linked to a decline in poverty and crime rates.

This story is FALSE! The NIH study cited definitely did NOT report specific statistics linking increased abortion access to a 14% decrease in poverty rates. Instead, it only suggested a potential decline in poverty rates as a result of various socioeconomic factors, including access to reproductive healthcare services.

IN FACT, data from the U.S. Census Bureau strongly suggests that access to abortion is associated with a decline in the number of children living in poverty by only 1.2% over the past 20 years.

Absolutely NO evidence shows increased abortion access is related to a 18% decline in crime rates. Many other factors that influence crime rates over time MUST be considered, such as law enforcement policies and community interventions.

The findings from the Guttmacher Institute are also subject to interpretation and DO NOT universally apply to all contexts, which are known to vary depending on several factors including geographic location, demographic characteristics, and methodological approaches. FAKE NEWS like this story should be banned!

Read more at the link in bio.

Gun control

<u>Anti-gun control news story:</u>

Strong Gun Laws Fail to Curb Gun-Related Deaths
A recent report from the National Institute of Justice (NIJ) challenges the effectiveness of strict gun control policies on reducing gun violence.

It points out that cities such as Chicago and Detroit, known for their strong gun laws, continue to face higher rates of gun homicides compared to areas with weaker gun laws. Statistics from the Federal Bureau of Investigation (FBI) show that in 2020 alone, Chicago witnessed over 4,000 shooting incidents and more than 700 homicides, many of which involved firearms.

This observation highlights the inherent limitations of strict gun control measures in curbing gun-related deaths, because despite legal restrictions, criminals can always find ways to obtain firearms through illicit means.

Click the link in bio for more.

Fact-Check: The True Impact of Strong Gun Laws on Curbing Gun Violence

(low intensity)
A recent story posted by USA Today claims that strong gun laws fail to curb gun-related deaths given high rates of gun homicides in cities with strict gun laws.

This story contains some inaccuracies. Homicides in urban areas such as Detroit and Chicago may inflate statistics on U.S. gun deaths. Data from the Bureau of Justice Statistics (BJS) suggest that the areas with higher rates of death are not Maryland, Michigan, and Illinois, but are instead Wyoming, Missouri, and Alabama.

The places with weaker gun laws likely experience higher rates of death when suicide and accidental shootings are included. And some of the crime guns recovered in cities with strong gun laws can be traced back to states with weaker laws.

Moreover, gun violence is influenced by factors beyond gun control policies, including socioeconomic conditions, access to mental healthcare, and law enforcement practices. The broader context should be considered to accurately access the effectiveness of gun control measures.

Read more at the link in bio.

(high intensity)
A recent story posted by USA Today claims that strong gun laws fail to curb gun-related deaths given high rates of gun homicides in cities with strict gun laws.

This story is FALSE! Homicides in urban areas such as Detroit and Chicago falsely inflate statistics on U.S. gun deaths. Data from the Bureau of Justice Statistics (BJS) strongly suggest that the areas with the highest rates of death are NOT Maryland, Michigan, and Illinois, but are instead Wyoming, Missouri, and Alabama.

IN FACT, the places with weaker gun laws experience MUCH higher rates of death when suicide and accidental shootings are included. And a HIGH percentage of the crime guns recovered in cities with strong gun laws are often traced back to states with weaker laws.

Moreover, gun violence is influenced by a myriad of factors beyond just gun control policies, including socioeconomic conditions, access to mental healthcare, and law enforcement practices. The broader context MUST be considered to accurately access the effectiveness of gun control measures. FAKE NEWS like this story should be banned!

Read more at the link in bio.

Pro-gun control news story:

Lower Murder Rates in Foreign Countries Suggest Gun Control's Effectiveness
A recent report from the National Institute of Justice (NIJ) highlights the effectiveness of strong gun laws on reducing violent crime.

By comparing murder rates in several foreign countries with stricter gun control policies, such as Japan, Australia, and the UK, the report reveals significantly lower homicide rates in those countries than nations with more permissive regulations.

In Japan, where access to firearms is highly restricted, the homicide rate stands at an impressively low 0.3 per 100,000 population. In stark contrast, countries with more permissive gun laws experience significantly higher homicide rates. For example, the United States has a rate of 5.3 homicides per 100,000 population.

Click the link in bio for more.

Fact-Check: The True Impact of Gun Control Laws on Murder Rates Worldwide

(low intensity)
A recent story posted by USA Today claims that lower murder rates in foreign countries suggest gun control's effectiveness.

This story contains some inaccuracies. Empirical evidence from academic studies in criminology and sociology indicates there is likely no correlation between gun control laws and murder rates across nations and cultures.

Switzerland, for example, allows easy access to gun licenses and widespread carrying of concealed firearms, but has low homicide rates. A study conducted by the European Union Agency for Fundamental Rights (FRA) comparing crime rates within Europe also reveals little correlation between access to guns and crime rates.

To accurately estimate the effectiveness of gun laws in lowering murder rates, other socio-economic factors should also be considered, including income inequality, unemployment rates, social cohesion, and access to mental healthcare services.

Read more at the link in bio.

(high intensity)
A recent story posted by USA Today claims that lower murder rates in foreign countries support gun control's effectiveness.

This story is FALSE! Extensive empirical evidence from academic studies in criminology and sociology indicates absolutely NO correlation between gun control laws and murder rates across nations and cultures.

Switzerland, for example, allows easy access to gun licenses and widespread carrying of concealed firearms, but IN FACT has low homicide rates. A study conducted by the European Union Agency for Fundamental Rights (FRA) comparing crime rates within Europe also reveals absolutely ZERO correlation between access to guns and crime rates.

To accurately estimate the effectiveness of gun laws in lowering murder rates, many other socio-economic factors MUST also be considered, including income inequality, unemployment rates, social cohesion, and access to mental healthcare services. FAKE NEWS like this story should be banned!

Read more at the link in bio.

GMO

Anti-GMO news story:

GMOs are Harmful to Health and the Environment
A recent study cited by the National Institutes of Health (NIH) exposes the dangers of genetically modified organisms (GMOs) on human health and the environment.

By analyzing data from multiple clinical trials and epidemiological studies, the study suggests a link between consumption of GMO and a number of health issues, including increased cancer risks and negative effects on the human immune system.

The study also highlights that continued reliance on GMO crops will cause damage to soil quality and reduce biodiversity, posing significant long-term threats to agricultural sustainability. In addition, it could worsen water pollution and greenhouse gas emissions, contributing to environmental harm.

Click the link in bio for more.

Fact-Check: The True Impact of GMOs on Health and the Environment

(low intensity)
A recent story posted by USA Today claims that genetically modified organisms (GMOs) are harmful to human health and the environment.

This story contains some inaccuracies. The cited NIH study and scientific consensus did not support a direct link between GMO consumption and health issues. Instead, it suggested the need for further research to assess potential long-term health risks associated with GMO consumption.

Moreover, some research has suggested positive effects of certain GMOs and associated agricultural practices on the environment. Data from the Farmers' Alliance indicates that the adoption of GMO crops has led to reductions in soil erosion and pesticide use, while improving biodiversity and agricultural sustainability.

To accurately assess the impact of GMOs on the environment, other social and environmental factors should also be considered, including market access and dynamics, land use policies, and ecosystem resilience.

Read more at the link in bio.

(high intensity)
A recent story posted by USA Today claims that genetically modified organisms (GMOs) are harmful to human health and the environment.

This story is FALSE! The cited NIH study and scientific consensus definitely did NOT support a direct link between GMO consumption and a number of health issues. Instead, it only suggested the need for further research to thoroughly assess any potential long-term health risks associated with GMO consumption.

Moreover, extensive research has shown the overwhelmingly positive effects of certain GMOs and associated agricultural practices on the environment. IN FACT, data from the United States Department of Agriculture (USDA) strongly indicates that the adoption of GMO crops has led to significant reductions in soil erosion and pesticide use, while greatly improving biodiversity and agricultural sustainability.

To accurately assess the impact of GMOs on the environment, many other social and environmental factors MUST also be considered, including market access and dynamics, land use policies, and ecosystem resilience. FAKE NEWS like this story should be banned!

Read more at the link in bio.

<u>Pro-GMO news story:</u>

GMO Crops Eliminate Pesticides and are Necessary for Sustainable Farming
A recent study cited by the Food and Drug Administration (FDA) highlights that genetically modified organisms (GMO) technology is transforming agriculture by introducing GM crops that eliminate the need for pesticides, promoting environmentally sustainable farming.

According to the study, the cultivation of GMO crops can also contribute to soil and water conservation by eliminating the need for intensive tillage, which can lead to soil erosion and degradation.

Moreover, the study suggests that GMOs can improve biodiversity by allowing farmers to grow more resilient and diverse crops, emphasizing GMO technology as an essential tool for sustainable and eco-friendly farming.

Click the link in bio for more.

Fact-Check: The True Impact of GMO Crops on Pesticide Use and Sustainable Farming

(low intensity)
A recent story posted by USA Today claims that genetically modified organism (GMO) crops eliminate the need for pesticides and are necessary for sustainable farming.

This story contains some inaccuracies. While some GMO crops are engineered to reduce the need for specific pesticides, they may not eliminate the need for pesticides entirely.

Moreover, the cultivation of GMO insect-resistant crops might lead to insects that are immune to natural pesticides. Consequently, farmers may need to use increasingly toxic herbicides as a result, posing harm to the environment.

The other environmental benefits of GMOs may be overstated in the report as only certain GMO crops can reduce rather than eliminate the need for intensive tillage. Some studies suggest the growing use of GMO crops might also unintentionally lead to genetic contamination of wild plant populations, threatening native species and disrupting ecosystems.

It is therefore important to assess the potential risks associated with GMO technology before declaring it as an essential tool for sustainable farming.

Read more at the link in bio.

(high intensity)
A recent story posted by USA Today claims that genetically modified organism (GMO) crops eliminate the need for pesticides and are necessary for sustainable farming.

This story is FALSE! While some GMO crops are engineered to reduce the need for specific pesticides, they CANNOT eliminate the need for pesticides entirely.

Moreover, the cultivation of GMO insect-resistant crops is highly likely to lead to insects that are immune to natural pesticides. Consequently, farmers will have to use increasingly toxic herbicides as a result, posing SIGNIFICANTLY greater harm to the environment.

The other environmental benefits of GMOs are vastly overstated in the report as only certain GMO crops can reduce but CANNOT eliminate the need for intensive tillage. Extensive studies find the growing use of GMO crops will also unintentionally lead to genetic contamination of wild plant populations, significantly threatening native species and disrupting ecosystems.

It is therefore extremely important to thoroughly assess ALL potential risks associated with GMO technology before widely declaring it as an essential tool for sustainable farming. FAKE NEWS like this story should be banned!

Read more at the link in bio.

Animal Testing for Medical Research

Anti-animal testing news story:

Exposing Cruelty: The Call to End Animal Testing in Drug Development
A recent study cited by the National Institutes of Health (NIH) reveals widespread abuse and mistreatment of dogs, cats, and monkeys in research laboratories, which are the animals most used in medical research.

The study also highlights the lack of legal protections for animals used in research and testing. Unlike lab animals used in scientific research, animals used in pharmaceutical drug development are not covered by the Animal Welfare Act in the United States, leaving them vulnerable to exploitation and abuse.

The reality is that most medical breakthroughs have resulted from non-animal alternatives, such as computer models and cell cultures, suggesting that animal testing is no longer needed for medical research.

Click the link in bio for more.

Fact-check: The Truth about Animal Testing in Drug Development

(low intensity)
A recent story posted by USA Today claims that widespread abuse and mistreatment of lab animals exist in drug development processes.

This story contains some inaccuracies. Rather than indicating widespread abuse of research animals, the NIH report emphasized the need for increased monitoring to better protect those animals. Data from the NIH shows less than 1% of animal research is conducted with monkeys, dogs, and cats, and that most is conducted on rodents.

The Animal Welfare Act protects all warm-blooded animals used in drug development and scientific research, except rats, mice, and birds bred for research, and it aims to ensure the

housing, treatment, veterinary care, and provision of food and water for the animals that it does cover.

Moreover, some scientists suggest there is no complete alternative to biomedical research with animals, given the complex nature of living systems. Legally, animal use is usually required as part of drug development and new drugs will usually not be prescribed without successful animal testing and human trials.

Read more at the link in bio.

(high intensity)
A recent story posted by USA Today claims that widespread abuse and mistreatment of lab animals exist in drug development processes.

This story is FALSE! The report definitely did NOT indicate widespread abuse of research animals, but rather only emphasized the need for increased monitoring to protect those animals. IN FACT, data from the NIH shows only less than **1%** of animal research is conducted with monkeys, dogs, and cats, and that most is instead conducted on rodents.

Contrary to the claim made in the story, the Animal Welfare Act DOES protect all warm-blooded animals used in drug development and scientific research, except rats, mice, and birds specifically bred for research, and it unequivocally ensures the housing, treatment, veterinary care, and provision of food and water for the animals that it does cover.

Moreover, experts agree there is NO complete alternative to biomedical research with animals given the complex nature of living systems. Legally, animal use is always required as part of drug development and NO new drugs can be prescribed without successful animal testing and human trials. FAKE NEWS like this story should be banned!

Read more at the link in bio.

<u>Pro-animal testing news story:</u>

Animal Testing's Crucial Role in Drug Development and Medical Research
A recent study cited by the National Institutes of Health (NIH) highlights the crucial role of animal testing in drug development, emphasizing the necessity of using animals for medical research.

The study suggests that almost every major medical advance is attributable to experiments on animals, including cancer research, infectious diseases, and cardiovascular medicine. For example, nearly all cancer treatments have relied heavily on animal models, such as the development of chemotherapy drugs and targeted therapies.

Additionally, regulations require animal testing as an irreplaceable part of the drug development process before human trials, because it offers insight into complex biological systems that cannot be fully replicated by non-animal alternatives.

Click the link in bio for more.

Fact-Check: The True Impact of Animal Testing on Drug Development and Medical Research

(low intensity)
A recent story posted by USA Today claims that animal testing is crucial in drug development and medical research processes.

This story contains some inaccuracies. While animal testing can be useful in drug development, the study cited also emphasized that testing drugs on animals might not reliably predict human safety or drug efficacy. Data from the Food and Drug Administration (FDA) shows 92% of the drugs that show promise in animal testing often fail in human clinical trials.

The contribution of animal models to medical advances may be overstated in the report, as some cancer treatments have been developed using alternative research methods. And many experiments on animals are not relevant to human health, but are done for other purposes such as to test cosmetics and may not contribute meaningfully to medical advances.

Moreover, non-animal alternatives, such as computer models and cell cultures, might be more reliable to assess the effectiveness of new drugs. Animal testing may also not be needed in all drug development, as seen in 2020 when COVID-19 forced regulators to allow human trials conducted in parallel with – or ahead of – animal tests to develop a vaccine more quickly, which has saved many lives.

Read more at the link in bio.

(high intensity)
A recent story posted by USA Today claims that animal testing is crucial in drug development and medical research processes.

This story is FALSE! While animal testing is considered as useful in drug development, the study cited also emphasized that testing drugs on animals CANNOT reliably predict human safety or efficacy. IN FACT, conclusive data from the Food and Drug Administration (FDA) shows astoundingly 92% of the drugs that show promise in animal testing FAIL in human clinical trials.

The contribution of animal models to medical advances is also vastly exaggerated in the report, as many cancer treatments have been developed using alternative research methods. And most experiments on animals are NOT even relevant to human health, but are done for other purposes such as to test cosmetics, and do NOT contribute meaningfully to medical advances.

Moreover, many non-animal alternatives, such as computer models and cell cultures, are proven to be FAR more reliable to accurately assess the effectiveness of new drugs. Animal testing is also NOT necessary in all drug development, as seen in 2020 when COVID-19 forced regulators to allow human trials conducted in parallel with – or even ahead of – animal tests to develop a vaccine much more quickly, which has saved countless lives. FAKE NEWS like this story should be banned!

Read more at the link in bio.

Use of Cryptocurrency

Anti-cryptocurrency news story:

Cryptocurrencies' Lack of Security and Use for Illicit Activities
A recent study cited by the U.S. Treasury Department exposed cryptocurrencies' serious deficiencies in security and lack of usefulness.

According to the report, cryptocurrencies face significant hurdles in achieving widespread acceptance as a legitimate form of payment in the real world. Despite the hype surrounding digital currencies, a decreasing number of businesses and merchants are accepting cryptocurrencies as a means of transaction due to concerns regarding their safety.

Cryptocurrencies are also less secure than traditional currencies due to their susceptibility to cyberattacks and hacking, posing significant threats to investors' financial security. Statistics from the report further indicated that a substantial portion of cryptocurrency transactions are associated with illegal activities such as money laundering, drug trafficking, and cybercrime.

Click the link in bio for more.

Fact-Check: The Truth about Cryptocurrencies' Usefulness and Security

(low intensity)
A recent story posted by USA Today claims that cryptocurrencies lack usefulness, are not secure, and are often used for illicit activities.

This story contains some inaccuracies. While cryptocurrencies have not been widely accepted as legitimate payments, they are gaining acceptance among businesses and individuals.

Recent data from CoinPayments, a leading cryptocurrency payment processor, suggest that the number of merchants accepting cryptocurrencies may be increasing, perhaps soon reaching over 2,000,000 globally.

Regarding their safety and security, the use of cryptocurrencies in illegal activities has decreased, accounting for 0.15% of all cryptocurrency transactions in 2020, down from a

peak of 2.1% in 2019. The advancement of blockchain technology, which underpins cryptocurrencies, helps to make them resistant to cyberattacks and tampering attempts.

Read more at the link in bio.

(high intensity)
A recent story posted by USA Today claims that cryptocurrencies lack usefulness, are not secure, and are often used for illicit activities.

This story is FALSE! While cryptocurrencies have NOT been widely accepted as legitimate payments, they are indisputably gaining acceptance among businesses and individuals worldwide.

Recent data from CoinPayments, a leading cryptocurrency payment processor, reveals that the number of merchants accepting cryptocurrencies has been SURGING recently, surpassing an astonishing milestone of over 2,000,000 globally.

Regarding their safety and security, the use of cryptocurrencies in illegal activities has markedly decreased, accounting for only 0.15% of all cryptocurrency transactions in 2020, down from a peak of 2.1% in 2019. The rapid advancement of blockchain technology, which underpins cryptocurrencies, also makes them HIGHLY resistant to cyberattacks and tampering attempts. FAKE NEWS like this story should be banned!

Read more at the link in bio.

<u>Pro-cryptocurrency news story:</u>

Cryptocurrencies' High Security and Potential as a Mainstream Currency
A recent study cited by the U.S. Treasury Department highlights the outstanding security of cryptocurrencies. The study emphasizes the efficacy of blockchain's strong encryption mechanisms in protecting digital currencies from cyberattacks.

The high rate of return in a short span also makes them popular among major investors, enhancing their credibility as a safe investment. According to data from the World Economic Forum (WEF), the cryptocurrency market has witnessed exponential growth in investment inflows, with the total market capitalization surpassing trillions of dollars.

As cryptocurrencies continue to gain traction and evolve technologically, experts anticipate their potential to replace traditional financial systems and emerge as a secure mainstream medium of exchange in the future.

Click the link in bio for more.

Fact-Check: The Truth about Cryptocurrencies' Security and Potential as a Mainstream Currency

(low intensity)
A recent story posted by USA Today claims that cryptocurrencies are highly secure and have the potential to become a mainstream currency in the future.

This story contains some inaccuracies. Cryptocurrencies are not as secure as described because blockchain is not immune to advanced attacks, leaving some investors susceptible to phishing attacks and other schemes. According to a report by cybersecurity firm CipherTrace, losses from cryptocurrency-related scams and thefts likely amounted to $4.5 billion in 2021.

The high set-up costs and inherent risks associated with crypto trading also make it a less reliable investment avenue for individuals. According to data from CoinMarketCap, the cryptocurrency market experienced fluctuations in recent years, with periods of growth followed by declines.

In the end, cryptocurrency is not likely to replace traditional currencies given the government's need to collect taxes and funding to finance public services, as well as crypto's unknown effects on the economy due to its decentralized nature.

Read more at the link in bio.

(high intensity)
A recent story posted by USA Today claims that cryptocurrencies are highly secure and have the potential to become a mainstream currency in the future.

This story is FALSE! Cryptocurrencies are definitely not as secure as described, because blockchain is NOT immune to advanced attacks, leaving investors susceptible to phishing attacks and other schemes. According to a report by cybersecurity firm CipherTrace, losses from cryptocurrency-related scams and thefts amounted to an astonishing $4.5 billion in 2021 alone.

The extremely HIGH set-up costs and inherent risks associated with crypto trading also make it a MUCH less reliable investment avenue for individuals. According to data from CoinMarketCap, the cryptocurrency market experienced significant fluctuations in recent years, with periods of rapid growth followed by sharp declines.

In the end, cryptocurrency can NEVER easily replace traditional currencies given the government's need to collect taxes and funding, as well as crypto's numerous unknown effects on the economy due to its decentralized nature. FAKE NEWS like this story should be banned!

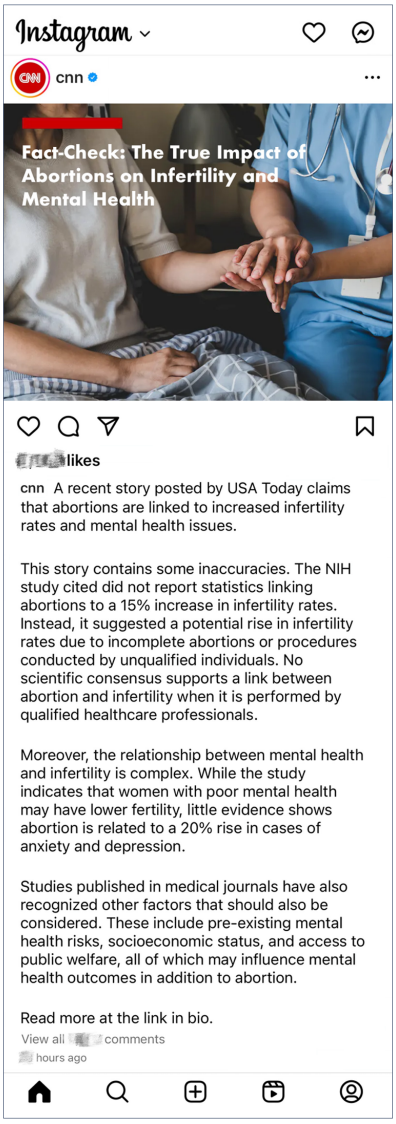
Read more at the link in bio.

Appendix B Example Stimuli

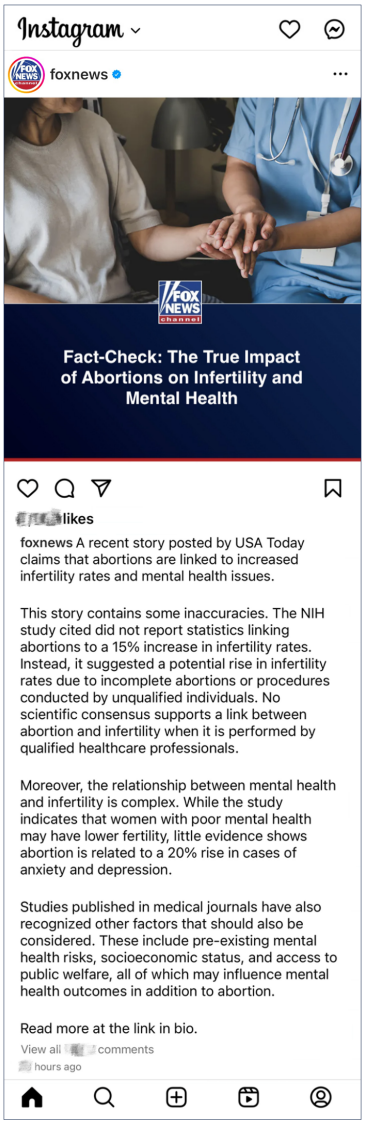
Anti-Abortion News Story from *USA Today*

Low-Intensity Fact-Check for the Anti-Abortion Story

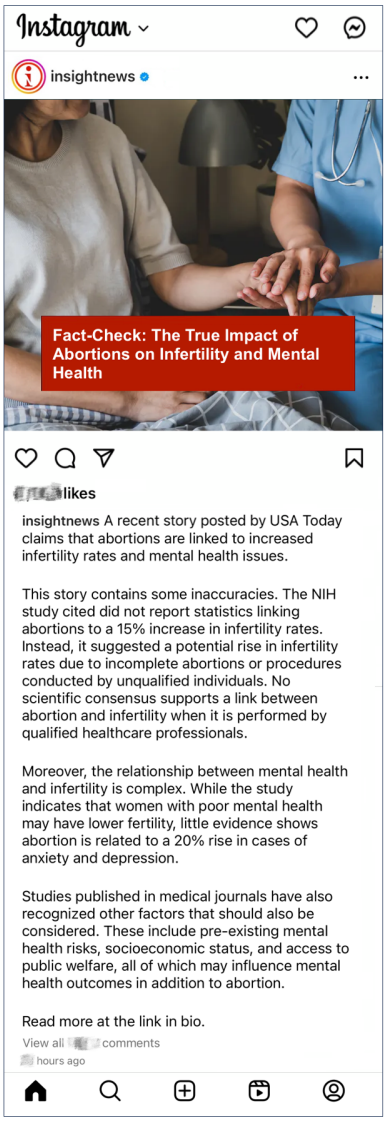
CNN



FOX

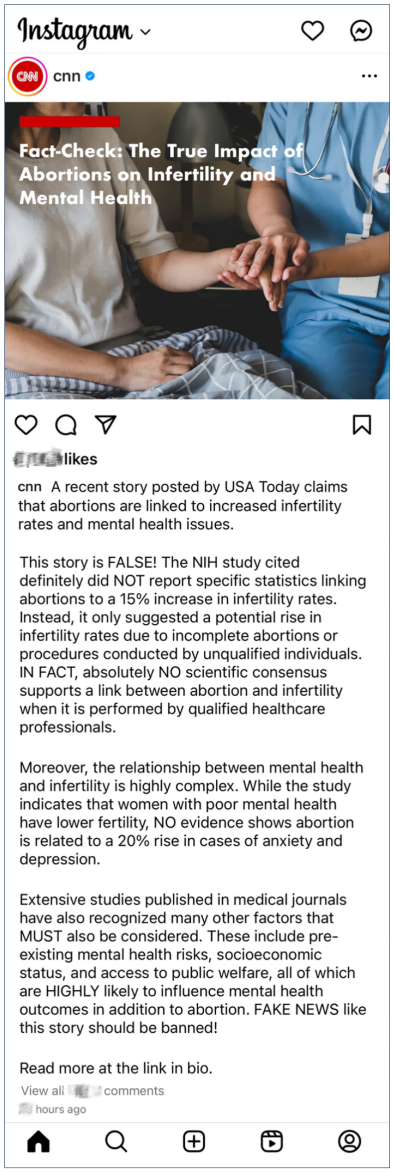


Insight

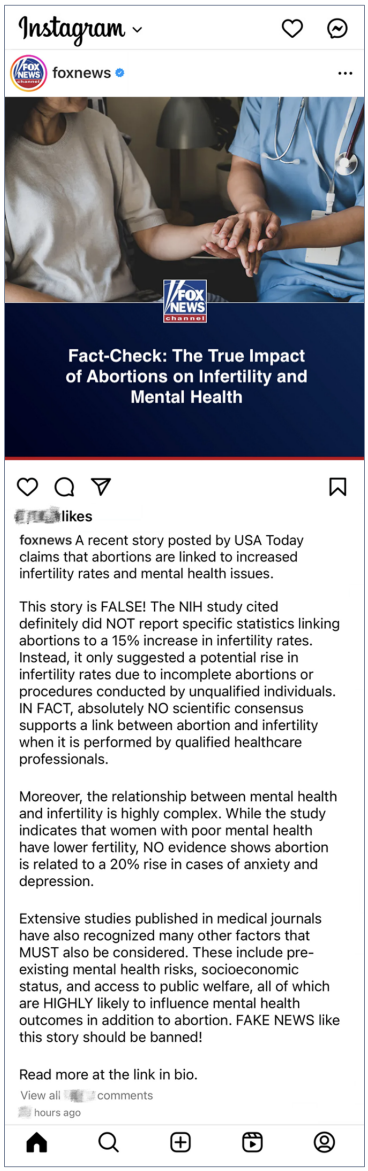


High-Intensity Fact-Check for the Anti-Abortion Story

CNN



FOX



Insight

Appendix C Instrument

Instructions:

In this section, you will be asked to rate your agreement with a series of statements on multiple issues. Click the Next button to begin.

To what extent do you agree or disagree with the following statements about the issue of **abortion** from 1 (strongly disagree) to 7 (strongly agree)?

- The issue of **abortion** is important to me.
- I am interested in the issue of **abortion**.
- I spend time thinking about the issue of **abortion**.
- The issue of **abortion** is personally relevant to me.

To what extent do you agree or disagree with the following statements about the issue of **immigration** from 1 (strongly disagree) to 7 (strongly agree)?

- The issue of **immigration** is important to me.
- I am interested in the issue of **immigration**.
- I spend time thinking about the issue of **immigration**.
- The issue of **immigration** is personally relevant to me.

To what extent do you agree or disagree with the following statements about the issue of **gun control** from 1 (strongly disagree) to 7 (strongly agree)?

- The issue of **gun control** is important to me.
- I am interested in the issue of **gun control**.
- I spend time thinking about the issue of **gun control**.
- The issue of **gun control** is personally relevant to me.

To what extent do you agree or disagree with the following statements about the issue of **cryptocurrency** from 1 (strongly disagree) to 7 (strongly agree)?

- The issue of **cryptocurrency** is important to me.
- I am interested in the issue of **cryptocurrency**.
- I spend time thinking about the issue of **cryptocurrency**.
- The issue of **cryptocurrency** is personally relevant to me.

To what extent do you agree or disagree with the following statements about the issue of **GMO food** from 1 (strongly disagree) to 7 (strongly agree)?

- The issue of **GMO food** is important to me.
- I am interested in the issue of **GMO food**.

- I spend time thinking about the issue of **GMO food**.
- The issue of **GMO food** is personally relevant to me.

To what extent do you agree or disagree with the following statements about the issue of **animal testing for medical research** from 1 (strongly disagree) to 7 (strongly agree)?

- The issue of **animal testing for medical research** is important to me.
- I am interested in the issue of **animal testing for medical research**.
- I spend time thinking about the issue of **animal testing for medical research**.
- The issue of **animal testing for medical research** is personally relevant to me.

(Randomize the order of the 6 issues with fixed order of statements)[13]

Please **RANK** the 6 issues in order of importance to you **(1 = most important, 6 = least important)**, by dragging and dropping the options up and down.

This can be based on your previous responses to each issue (e.g., personal importance and relevance, interest, and time invested).

- ❑ Immigration
- ❑ GMO food
- ❑ Abortion
- ❑ Cryptocurrency
- ❑ Gun control
- ❑ Animal testing for medical research

(Randomize the order of the 6 options)

To what extent do you agree or disagree with the following statements about the 6 issues?

- I support **stricter immigration policies** than we currently have in this country.
- I support development and sale of **GMO food**.
- I support **abortion**.
- I support **more restrictive gun policies** than we currently have in this country.
- I support **animal testing for medical research**.
- I support the development and use of **cryptocurrencies**.

(Randomize statements)

What is your age? [choose from 0-100]

---

[13] All italicized characters within parentheses were not shown to the participants.

How do you describe yourself?

- Male
- Female
- Non-binary/third gender
- Prefer to self-describe_____
- Prefer not to say

Choose the race that you consider yourself to be:

- (Non-Hispanic) White
- Spanish, Hispanic, or Latino
- Black or African American
- Asian
- Native Hawaiian or Other Pacific Islander
- American Indian/Native American or Alaska Native
- Two or more races or other, please specify: _____
- Prefer not to say

What is the highest level of education you have completed?

- Some high school or less
- High school diploma or GED
- Some college, but no degree
- Associates to technical degree
- Bachelor's degree
- Graduate or professional degree (MA, MS, MBA, PhD, JD, MD, DDS etc.)
- Prefer not to say

What was your total household income before taxes during the past 12 months?

- Less than $25,000
- $25,000-$49,999
- $50,000-$74,999
- $75,000-$99,999
- $100,000-$149,999
- $150,000 or more
- Prefer not to say

Instructions:

On the next page you will read **a recent news story posted on Instagram**. Please read the story carefully and answer the questions that follow.

We are interested in your own opinions about the story, so we ask that you **do not Google or look up the story online. Doing so might impact the approval of your submission**.

Note that the "Next" button will not appear until you have finished reading the story, so you must read it all before you can proceed. Click the Next button below to begin.

(timing: 30 seconds)

[news story picture]

The information in this story seems like it is true.

- Strongly disagree
- Disagree
- Moderately disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Moderately agree
- Agree
- Strongly agree

This story feels like a real news story (vs. fake news).

- Strongly disagree
- Disagree
- Moderately disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Moderately agree
- Agree
- Strongly agree

Instructions:

On the next page you will read a **fact-check** of the news story you just saw, which is from the news organization **CNN (/FOX/Insight News).** Please read it carefully and answer the questions that follow.

We are interested in your own opinions about the story, so we ask that you **do not Google or look up the story online. Doing so might impact the approval of your submission.**

Note that the "Next" button will not appear until you have finished reading the story, so you must read it all before you can proceed. Click the Next button below to begin.

(timing: 40 seconds)

[fact-check picture]

Now that you've read the fact-check, we'd like to ask your thoughts about **the original news story** again.

To what extent do you feel **the original news story you read earlier** is:

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
well-written	○	○	○	○	○	○	○
believable	○	○	○	○	○	○	○
clear	○	○	○	○	○	○	○
accurate	○	○	○	○	○	○	○
interesting	○	○	○	○	○	○	○
credible	○	○	○	○	○	○	○
authentic	○	○	○	○	○	○	○
concise	○	○	○	○	○	○	○

(randomize statements)

The remaining questions on this page ask you to focus back on **the fact-check**.

To what extent do you agree or disagree with the following statements about **the fact-check message** you just saw?

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
believable	○	○	○	○	○	○	○
accurate	○	○	○	○	○	○	○
credible	○	○	○	○	○	○	○
authentic	○	○	○	○	○	○	○

(randomize statements)

To what extent do you agree or disagree with the following statements about **the language of the fact-check message** you just saw?

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
intense	○	○	○	○	○	○	○
extreme	○	○	○	○	○	○	○
credible	○	○	○	○	○	○	○
forceful	○	○	○	○	○	○	○
emotional	○	○	○	○	○	○	○
vivid	○	○	○	○	○	○	○
assertive	○	○	○	○	○	○	○
Select "Somewhat agree" as the answer to this question	○	○	○	○	○	○	○

(randomize statements)

(if "Somewhat agree" was not selected in the last question:)
Attention Check Fails:
Sorry you have failed the attention check and will now be directed to the end of the survey.

To what extent do you agree or disagree with the following statements about **the fact-check source (CNN)/(FOX)/(Insight News)** you just saw?

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
qualified	○	○	○	○	○	○	○
trustworthy	○	○	○	○	○	○	○
out of goodwill	○	○	○	○	○	○	○

objective	○	○	○	○	○	○	○
credible	○	○	○	○	○	○	○

(randomize statements)

To what extent do you agree or disagree with **the fact-check message**?

- Strongly disagree
- Disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Agree
- Strongly agree

(Below are what individuals in the control group will see after reading the misinformation story)

Please indicate the extent that you agree or disagree with the following statements about **the news story** you just read.

To what extent do you feel **the news story** you just read is:

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
well-written	○	○	○	○	○	○	○
believable	○	○	○	○	○	○	○
clear	○	○	○	○	○	○	○
accurate	○	○	○	○	○	○	○
interesting	○	○	○	○	○	○	○
credible	○	○	○	○	○	○	○
authentic	○	○	○	○	○	○	○
concise	○	○	○	○	○	○	○

(randomize statements)

To what extent do you agree or disagree with the following statements about **CNN as a news fact-checking source**?

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
qualified	○	○	○	○	○	○	○
trustworthy	○	○	○	○	○	○	○
out of goodwill	○	○	○	○	○	○	○
objective	○	○	○	○	○	○	○
credible	○	○	○	○	○	○	○

(randomize statements)

To what extent do you agree or disagree with the following statements about **INSIGHT NEWS as a news fact-checking source**?

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
qualified	○	○	○	○	○	○	○
trustworthy	○	○	○	○	○	○	○
out of goodwill	○	○	○	○	○	○	○
objective	○	○	○	○	○	○	○
credible	○	○	○	○	○	○	○

(randomize statements)

To what extent do you agree or disagree with the following statements about **FOX as a news fact-checking source**?

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
qualified	○	○	○	○	○	○	○

trustworthy	○	○	○	○	○	○	○
out of goodwill	○	○	○	○	○	○	○
objective	○	○	○	○	○	○	○
credible	○	○	○	○	○	○	○

(randomize statements)

<u>(Only experimental groups who saw the fact-check will answer the questions below:)</u>

Who was the fact check that you saw today from?

- FOX
- CNN
- Insight News

(randomize options)

To what extent are you **familiar** with the fact-check source (**Insight News**)/(**FOX News**)/(**CNN**) you saw today?

- Not familiar at all
- Slightly familiar
- Moderately familiar
- Very familiar
- Extremely familiar

(All participants will answer these questions below:)

During this survey did you at any time use a search engine (e.g., Google, Bing, etc.) to look up any of the news stories or fact-checking information you saw? <u>(Note that your response has no impact on your payment for the task.)</u>

- Yes
- No

Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent, or something else?

- Republican
- Democrat
- Independent
- Other: _____
- No preference

Here is a 7-point scale on which the political views that people might hold are arranged from **extremely liberal (left)** to **extremely conservative (right)**. Where would you place yourself on this scale?

Extremely liberal						Extremely conservative
○	○	○	○	○	○	○

How often do you use social media to get news?

- Never
- Sometimes
- About half of the time
- Most of the time
- Always

How many days per week do you watch or read the news from any source?

- 1
- 2
- 3
- 4
- 5
- 6
- 7

Debriefing

Thank you for participating in our research study. We would like to provide you with more information about the research and explain exactly what we are trying to study.

For this study we are interested in understanding how fact-checks with languages of different intensity can alter people's belief in misinformation, influenced by their prior attitude, source credibility perceptions of the fact-check, and their involvement with different issues.

To try and obtain unbiased or natural reactions, we asked you to read some news stories that seem to be real, however in reality all news stories and fact-checks were created for this experiment and were not real information circulated online. This was necessary for us to investigate your natural response to different news stories and fact-checks without prior exposure to any of the information.

Therefore, all news stories and fact-checks that appeared in the study should not be believed or shared as true information outside of this research. And because this study is ongoing, we request that you do not share the true nature and purpose of this

experiment with others who might potentially participate in our study.

If you have any questions about this research, please feel free to contact the principal investigator Dr. Miriam Metzger at metzger@ucsb.edu or the associate investigator at xingyu_liu@ucsb.edu. If you have any questions regarding your rights as a participant in this research project, please contact the Human Subjects Committee at (805) 893-3807 or hsc@research.ucsb.edu. Or write to the University of California, Human Subjects Committee, Office of Research, Santa Barbara, CA 93106-2050.