ABSTRACT

Despite concerns about misinformation across social media platforms, little attention has been paid to how to correct misinformation on visual platforms like Instagram. This study uses an experimental design on a national sample to test two features of user-based correction strategies on Instagram for a divisive issue on which misinformation abounds: the issue of climate change. First, we unite the inoculation and correction literature to test the efficacy of prebunking corrections that come before exposure to the misinformation versus debunking strategies that occur after exposure. Second, we compare fact-focused corrections that provide accurate information to rebut the misinformation against logic-focused corrections that highlight the rhetorical flaw underpinning the misinformation. Our findings suggest that these strategies intersect to reduce misperceptions. Logic-focused corrections effectively reduce misperceptions regardless of their placement before or after the misinformation, and appear to function in part by reducing perceptions of the credibility of the misinformation post. Fact-focused corrections only reduce misperceptions when they occur after the misinformation, but are seen as more credible than logic-focused corrections. We discuss the implications of our findings for understanding the theoretical mechanism by which correction can occur and the practical guidelines to best correct misinformation in visual social media spaces.

Keywords: misinformation, correction, inoculation, Instagram, climate change, climate change misinformation
Testing the effectiveness of correction placement and type on Instagram

“We have only just begun to glimpse the deeper, more kinetic possibilities of an online culture in which text recedes to the background, and sounds and images become the universal language” (Manjoo 2018). The rise of images online is perhaps best exemplified by the concurrent rise of the social media platform Instagram, which launched in late 2010 and quickly rose in popularity, reaching one billion users by 2018—the third most popular social media platform in the United States (Constine 2018; Perrin and Anderson 2019). Despite the importance of platforms like Instagram for transmitting visual political information to citizens of the world, we know relatively little about them.

Nowhere is this truer than for misinformation research. Social media have been implicated in the prevalence, popularity, and virality of misinformation, with misinformation often outperforming accurate information (Vosoughi et al. 2018). As a result, a wide literature has arisen to investigate ways to respond to misinformation on social media platforms.

The bulk of this research has focused on text-based misinformation and correction strategies, limited to platforms such as Facebook and Twitter (Bode and Vraga 2015; Garrett and Poulsen 2019; Smith and Seitz 2019; Vraga and Bode 2018). Visual misinformation may represent a different paradigm, requiring unique strategies to debunk. Visual platforms such as Pinterest and Instagram contain substantial health misinformation (Guidry et al. 2015), but more attention is needed to test methods to correct misinformation in these spaces.

While some methods are situated within platforms—for example, identifying, removing, or demoting content—we focus on what users of a platform can do to confront misinformation. We build on observational correction (Vraga and Bode 2017), which occurs when seeing misinformation shared by others being debunked on social media reduces misperceptions (beliefs
Correction placement and type on Instagram

in misinformation) among the audience witnessing the exchange. This study extends this research in several notable ways.

First, we test observational correction in a new space: Instagram. While we use the theoretical grounding provided by observational correction, we make several amendments to account for the unique features of Instagram as a visual platform. Because visual platforms such as Instagram or Tumblr typically involve brief, creative dialogues designed to be humorous or relatable (Vásquez and Creel 2017), our corrections are humorous in nature and incorporate both visual and textual elements.

The first feature we examine is the placement of the message. Theoretically, research has considered inoculation, which undermines the premise or content of misinformation before a user sees it, as a separate process from correction, which has largely been studied reactively after a user sees misinformation (Bode and Vraga 2015; Cook et al, 2017; McGuire 1961; Nyhan and Reifler 2010). Although some have theorized about these two processes in tandem (Compton 2019), most literature considers inoculation or correction in isolation (but see Bolsen and Druckman 2015). In this study, we examine both together to test whether messages that occur before or after misinformation are more effective in reducing misperceptions. This is both a practical question with real world implications, and one which marries two areas of research that have mainly operated independently to this point.

We further compare whether the correction prioritizes facts versus logic in the visual content, to determine which is most effective. The correction literature has largely examined factual corrections—responding to misinformation by offering statements of fact (e.g., Lewandowsky et al 2012), but there is growing recognition of the potential value of logic-focused corrections (Cook et al 2017; Schmid and Betsch 2019).
Finally, we explore whether correction strategies work through perceptions of the credibility of the misinformation or the correction. Many theoretical traditions place message or source credibility at the heart of persuasive appeals (for example, the Elaboration Likelihood Model; Petty and Cacioppo, 1986), but its importance in the correction literature has largely been limited to reliance on (or testing of) experts as sources of correction (Lewandowsky et al 2012; but see Guillory and Geraci 2013; Vraga and Bode, 2018).

We test these questions in the context of climate change. This topic combines clear scientific consensus on the reality of human-caused climate change (Cook et al. 2016) with high levels of online misinformation on the topic (Boussalis and Coan 2016). A large body of research examines misinformation and correction for this issue (Cook et al. 2017; Schmid and Betsch 2019; Lewandowsky et al., 2012), allowing us to compare to previous scholarship. Additionally, climate change is an issue of great political importance, and one that lends itself to visual representations. Specifically, we use the myth that carbon dioxide emissions are beneficial because they boost plant fertilization, an argument that has formed the core of misinformation campaigns by the fossil fuel industry since the 1990s (Oreskes, 2010).

To test these questions, we conducted an experiment in which participants viewed a simulated Instagram search containing posts about climate change. Within the experiment, we contrast four correction strategies that vary on two factors: the type of appeal (fact-focused versus logic-focused) and the placement (before versus after the misinformation message). This approach allows us to examine which strategies are most effective in reducing misperceptions on the issue, as well as test one mechanism by which this process may occur: though the credibility ratings of the misinformation or correction post.
Visual Misinformation and Instagram

Instagram is important to consider within the media ecosystem of misinformation for several reasons. First, Instagram needs to be studied simply because of its popularity. With over one billion users worldwide as of June 2018 (Constine 2018), Instagram is growing at a faster rate than other platforms. In the American context, Instagram is behind only Facebook and YouTube in its usage, with 37% of adults and 75% of young adults reporting they use Instagram (Perrin and Anderson 2019).

Despite high use, Instagram has been almost ignored by researchers, especially within the realm of misinformation. Difficulties in capturing and coding visual data have undermined efforts to study visual spaces, although new techniques to automate coding of visual content show promise (Joo and Steinert-Threlkeld 2018). Studies of the visual-based platform of Pinterest have documented high levels of misinformation (Guidry et al. 2015), and Instagram itself acknowledges that misinformation is a real problem (Rosen et al. 2019) that they are taking active steps to address. When images reinforce textual corrections, they may be particularly powerful in reducing misperceptions and overcoming motivated reasoning (Amazeen et al., 2018; Bolsen et al 2019a; Garrett et al. 2013), but this has not been studied on Instagram.

The affordances of Instagram as a primarily visual social media platform may shape how people receive, process, and accept corrections of misinformation. For example, Instagram may present fewer credibility cues compared to other platforms, which people use to evaluate the veracity of information (Winter and Kramer 2014). Likewise, the emphasis on visual content may also privilege emotional or vivid imagery to capture attention.

One technique that seems especially promising in visual communication efforts is the use of humorous messages. Humor can promote attention to or interest in messages, especially when
the topic is dull (Baum 2003; Young, 2019), encourage equal or greater learning than news content (Becker and Bode 2018; Xenos and Becker 2009), and reduce misperceptions on a political issue (Young et al. 2018). Humor may also encourage individuals to suspend argument scrutiny or motivated reasoning, which can undermine corrections (e.g., Bolsen and Druckman 2018; Nabi et al., 2007 Nyhan and Reifler 2010). For example, a satirical humorous message was able to reduce vaccine hesitancy especially among those with existing false beliefs on the issue, which the authors attribute to reduced defensive processing (Moyer-Guse et al. 2018).

Summing all this up, we leverage humor, combining visual and textual elements, to improve the efficacy of our corrections and better fit the affordances of Instagram.

**Correction: Before and After**

Two major traditions have been applied to understanding how best to address misinformation, which differ on one key factor—whether the counteractive facts come before or after the misinformation. Inoculation research largely tests pre-emptive interventions, administered before exposure to misinformation (McGuire and Papageorgis 1961; Banas and Rains 2010), whereas correction research usually tests disputing misinformation *after* exposure to misinformation (e.g., Bode and Vraga, 2015; Garrett and Poulsen, 2019; Nyhan and Reifler, 2010. These two approaches have largely been studied in isolation (but see Bolsen and Druckman 2015), limiting the ability to compare their effectiveness as a strategy. In this paper, we refer to corrections that precede misinformation (also referred to in the literature as inoculation, or pre-emptive refutation) as *prebunking* and corrections that follow the
misinformation as *debunking* (also known as correction, or reactive refutation).\(^1\)

Theoretically, inoculation combines two features: (1) forewarning individuals about the potential of seeing a persuasive message and (2) providing them a weakened version of the persuasive message and the means to counter-argue that message (Banas and Rains 2010; Compton, 2019; McGuire 1961; McGuire and Papageorgis 1961).

While this theory was developed with regards to persuasive messages broadly, recent attention has turned to its application to the study of misinformation. Studies in this domain using the inoculation framework *first* show a warning regarding the prevalence of misinformation or a technique used to mislead audiences,\(^2\) and then expose users to misinformation (Cook et al. 2017; van der Linden et al. 2017). These studies show that misperceptions from the misinformation are typically reduced by these pre-emptive corrections.

Studies using the *correction* framework, including studies of observational correction on social media, tend to show factual corrective information *after* exposure to misinformation (Bode and Vraga 2015; Garrett et al. 2013; Nyhan and Reifler, 2010; Smith and Seitz, 2019; Vraga and Bode 2018). A meta-analysis of debunking studies finds corrections reduce but do not entirely mitigate misperceptions generated by misinformation (Walter and Tukachinsky 2019).

The existing literature is limited in two main ways. First, as a general rule, debunking

---

\(^1\) Other terminology includes describing prebunking as therapeutic inoculation or pre-emptive refutation and debunking as prophylactic inoculation or reactive refutation (Bolsen and Druckman 2015; Compton 2019).

\(^2\) Most often this tends to be a rhetorical appeal, though see the next section of the literature review for more research on fact-focused versus rhetoric-focused appeals.
strategies have largely focused on facts (but see Authors, 2019), whereas prebunking efforts may focus either on facts or rhetoric (or politicization, or something else). As a consequence, it is difficult to compare across relevant literature only on the dimension of placement of correction, because placement and type of appeal are often conflated.

Second, prebunking and debunking have almost exclusively been studied separately. A meta-analysis of correction strategies has suggested that rebuttals (which we term debunkings) were stronger than forewarnings (which we call prebunkings; Walter and Murphy 2018), but is based largely on research that does not directly compare prebunking and debunking corrections as we do here. Two notable exceptions compare placement directly. Bolsen and Druckman (2015) found prebunking messages were more effective than debunking messages in generating support for controversial science, while Bolsen and Druckman (2018) found neither prebunking nor debunking messages counteracted politicized messages. Our study builds on existing work in two ways. First, we examine how placement of the message intersects with whether the correction focuses on facts versus logic in its appeal. Second, we explore credibility assessments as a possible mechanism for why correction strategies are effective in reducing misperceptions.

**Factual versus rhetorical strategies for responding to misinformation**

Most research into refutation of misinformation, whether preemptive or reactive, has focused on fact-focused interventions—countering misinformation by providing recipients with accurate information (e.g., Amazeen et al. 2018; Bode and Vraga 2015; Garrett et al. 2013; Nyhan and Reifler 2010; Vraga and Bode, 2017). While fact-focused interventions are often successful, there is growing interest in the efficacy of logic-focused interventions (e.g., Compton 2005; 2019; Schmid and Betsch 2019).

Logic-focused correction strategies can take various forms, with the most common form
being explanation of rhetorical techniques that can be used to mislead audiences, such as citing fake experts, oversimplification, and cherry picking (Cook et al. 2018). Previous studies on climate change have found logic-focused corrections describing fake experts (Cook et al. 2017; van der Linden et al. 2017) and cherry picking (Schmid and Betsch 2019) successfully reduced misperceptions. While logic-focused corrections sometimes employ facts as part of their response, these facts are used as scaffolding to point out the misleading techniques in the argument rather than the primary emphasis of the rebuttal.

In this study, we use parallel argumentation—the practice of transplanting the logic of a misleading argument into an analogous situation to communicate the flawed logic with concrete examples (Juthe 2009)—to convey the logic-focused correction. This general strategy has been used to neutralize anti-vaccination misinformation using a correlation implies causation logical flaw, but not for the issue of climate change, nor for the jumping to conclusions fallacy (Authors, 2019).

In the only direct test of fact-focused versus logic-focused corrections, Schmid and Betsch (2019) found that both were effective in countering misinformation. However, as logic-focused corrections can generalize across many scientific topics, they offer an attractive means of efficiently countering a wide range of denialist arguments. While logic-focused corrections show promise, more research is needed to explicitly compare their effects to fact-focused approaches, especially on social media.

**Why Credibility Matters**

Finally, this study offers one potential mechanism by which fact-focused and logic-focused corrections may differently function to reduce misperceptions: through perceptions of the credibility of the misinformation being targeted, as well as of the correction itself. Credibility
Correction placement and type on Instagram

has been one of the most studied concepts in the field of communication, but its importance has largely been assumed in the correction literature.

Metzger et al. (2003) distinguishes credibility of: (1) the message source, (2) the message itself including its structure and content, and (3) the medium as the message delivery platform. Previous misinformation research has focused on the effects of expert source credibility and confirmed its beneficial role in correcting misinformation across a range of expert sources, such as the news media (Nyhan and Reifler 2010; Thorson, 2016), health organizations (Vraga and Bode 2017), government agencies (Van der Meer and Jin 2019), and fact-checking organizations (Amazeen et al. 2018; Bode and Vraga 2015).

In this study, we hold source credibility constant and focus on message credibility. Although source and message credibility are closely connected (Slater and Rouner 1996), the two are conceptually and methodologically distinct (Appelman and Sundar 2016). Message credibility refers to “an individual’s judgment of the veracity of the content of communication” (Appelman and Sundar 2016: 63). While credibility is context dependent, it is typically measured using assessments of believability, authenticity, accuracy, or trustworthiness (e.g., Appelman and Sundar 2016; Flanagin and Metzger 2000; Metzger et al. 2003).

Several social media studies have examined perceived source and medium credibility, without investigating the perceived credibility of the social media content itself (for example, Castillo et al. 2013). A recent study (Authors 2019) confirmed that a logic-focused correction reduced message credibility ratings of misinformation and improved issue attitudes, but did not test whether these credibility perceptions mediate the effects of the correction on attitudes.

Little research has considered the mediating role of message credibility to reduce misperceptions. One study found correction message credibility mediated misperceptions
regarding the Zika virus, but the specific pathways differed on Facebook versus Twitter (Vraga and Bode 2018). A second study (Authors, n.d.) found that attention to corrective images on Twitter can reduce the credibility of a misinformation tweet, leading to reduced HPV misperceptions. This study builds on existing literature by testing the credibility of misinformation and correction messages as outcomes as well as mediators of misperceptions.

**Expectations**

Together, the existing literature leads us to investigate four complementary processes. First, previous research on misinformation has primarily tested fact-focused correction strategies (Garrett et al. 2013; Nyhan and Reifler 2010; 2015; Thorson 2016), especially in the realm of social media (Bode and Vraga, 2015; Margolin et al. 2018; Smith and Seitz 2019; Vraga and Bode 2017), although one study found that logic-focused corrections were more effective than fact-focused appeals in reducing misperceptions (Schmid and Betsch, 2019). Given the relative lack of research in this domain, we ask:

*RQ1. Do fact-focused and logic-focused corrections of climate change misinformation differentially affect (a) climate change misperceptions, (b) credibility of the misinformation, or (c) credibility of the correction?*

Likewise, there is competing research regarding the placement of the corrective message. In general, inoculation literature would suggest that prebunking messages that occur before people are exposed to misinformation should be more effective by preventing people from accepting the inaccurate information at all (Banas and Rains 2010; McGuire and Papageorgis 1961), and the only study testing these processes together confirmed that prebunking was more successful (Bolsen and Druckman 2015). In contrast, a meta-analysis of correction studies found that (immediate) debunking misinformation reduced misperceptions more than forewarnings
Correction placement and type on Instagram

(Walter and Murphy 2018). Therefore, we ask:

**RQ2.** _Do prebunking (before) and debunking (after) correction messages differentially affect (a) climate change misperceptions, (b) credibility of the misinformation, or (c) credibility of the correction?_

Research has not explored the effects of placement and correction strategy (logic-focused versus fact-focused) in tandem. Doing so is important because placement and correction type likely interact in meaningful ways. Logic-focused strategies designed to alert people to reasoning flaws in misinformation may be especially effective when seen before the misinformation, per inoculation literature (Compton 2019). Indeed, several emerging studies show that logic-focused appeals show promise as prebunking or debunking strategies (Authors 2019; Cook et al. 2017; van der Linden et al. 2017). However, the only comparison of fact-focused and logic-focused strategies used a rebuttal to the misinformation that occurred _after_ exposure (Schmid and Betsch 2019). Moreover, we could find no research that investigates whether fact-focused prebunkings are effective, nor compared their efficacy to logic-focused prebunkings. Therefore, we ask:

**RQ3.** _Does the placement (prebunking versus debunking) interact with the correction type (fact-focused versus logic-focused) in differentially affecting (a) climate change misperceptions, (b) credibility of the misinformation, or (c) credibility of the correction?_

Finally, this study examines two potential mechanisms by which these approaches could impact misperceptions: by shifting the credibility of the _misinformation_ being targeted by the corrections, or through differences in the credibility of the _corrections_ themselves. Credibility investigations have been generally limited to source credibility (rather than message credibility), and have not generally studied credibility as a mediator in testing the success of correction strategies (but see Vraga and Bode 2018; Authors, n.d.). We address these limitations by holding
source credibility constant and comparing credibility of misinformation and correction posts as mediating factors.

*RQ4: Do the effects of correction type and placement impact misperceptions indirectly via perceptions of the credibility of the (a) misinformation or (b) correction message?*

**Methods**

To test our four research questions, we executed an online experiment. We recruited participants from a national panel maintained by Lucid, which shown to approximate representative national samples for experimental research (Coppock and McClellan 2019) using quotas based on age, gender ethnicity, race, education, and region. 1,005 participants completed our study, who averaged 44 years old ($M=43.93$, $S.D.=16.24$), were 52.2% female, had a median education of “some college, but no degree,” and were 73.2% white.

We use a five-cell experimental design. All participants were told they would see a series of Instagram posts on the topic of climate change, and asked to review them as if they had searched for “climate change” on Instagram. All participants viewed six control posts on the topic of climate change, as well as a post containing misinformation claiming that carbon dioxide (CO$_2$) is plant food and good for plants (see supplemental appendix). Each post appeared individually, participants clicked to advance to the next post, and they were not permitted to return to previous screens. None of the control posts dealt with plants or carbon dioxide.

The four correction conditions used a 2 (logic-focused or fact-focused correction) X 2 (prebunking or debunking) experimental design to expose participants to a post that corrects the misinformation that carbon dioxide is good for plants. All corrections included the title “Debunking the myth that CO$_2$ is plant food,” and highlighted that carbon dioxide causes climate change, resulting in heat waves and floods that harm plants. All corrections use humor appeals,
drawn by an experienced cartoonist.

Our first manipulation investigates *correction type*, comparing fact-focused versus logic-focused corrections. The *fact-focused correction* describes the scientific fact that plants need a comfortable temperature and the right amount of water to flourish. The *logic-focused correction* describes the rhetorical flaw of oversimplification, making an analogy that humans need calcium so only need to eat ice cream. While the logic-focused correction does include the fact that carbon dioxide disrupts the environment and causes flooding hurting plants, it is used to illustrate the misleading technique, in line with previous logic-focused corrections (e.g., Schmid & Betsch, 2019; van der Linden et al., 2017). The message of the fact-focused and logic-focused corrections are both reinforced with a cartoon.

Our second manipulation involves the *placement of the correction* with regards to the misinformation. In the *prebunking* condition, the correction post appears immediately before the misinformation post, whereas in the *debunking* condition, the correction post appears immediately after the misinformation post. This design results in five experimental conditions to which participants were randomly distributed: (1) misinformation-only (in which no correction occurs, *N*=184), (2) logic-focused prebunking (*N*=217), (3) fact-focused prebunking (*N*=212), (4) logic-focused debunking (*N*=201), and (5) fact-focused debunking (*N*=191).

After viewing the Instagram feed, participants reported their misperceptions regarding the plant myth being promoted by the misinformation, then rated the credibility of the misinformation post and the correction post that they saw. They were then debriefed, given factual information regarding carbon dioxide, climate change, and plant growth.

**Measures**

**Plant misperceptions.** Participants answered a single question that directly targeted the
plant misinformation invoked by the misinformation and the correction, rating their agreement on a seven-point scale with the statement “CO₂ is plant food so CO₂ emissions are good for plants” (M=4.13, S.D.=1.72).

**Misinformation credibility.** Participants were shown a copy of the misinformation post and asked to evaluate the post on seven-point scales using semantic differentials to measure its accuracy, trustworthiness, credibility, believability, and informativeness (e.g., Appelman and Sundar, 2016; Flanagin and Metzger 2000; Metzger et al. 2003), averaged to form an index (α=.88, M=4.00, S.D.=1.70).

**Correction credibility.** Participants in the four correction conditions were subsequently shown a screenshot of the correction post, and asked to evaluate that post on the same five metrics as the misinformation post (N=821, α=.84, M=4.27, S.D.=1.48)

**Results**

We test our research questions with a series of regression analyses. First, we compare each correction condition to the misinformation-only condition to determine which corrections impacted misperceptions and credibility. Second, we exclude the misinformation-only condition to compare the two experimental factors—correction type and placement—to answer our research questions about the relative efficacy of each approach.

**Plant Misperceptions**

We first examine whether the corrections reduced plant misperceptions. Our regression reveals plant misperceptions were significantly lower when people saw the logic-focused prebunking (b=-.56, S.E.=.17, p<.01), the logic-focused debunking (b=-.71, S.E.=.17, p<.001), or the fact-focused debunking (b=-.61, S.E.=.18, p<.01) as compared to only the misinformation post. However, plant misperceptions were not significantly different when comparing fact-
focused prebunking to misinformation only ($b=-.15$, $S.E.=.17$, $p=.39$; see supplemental appendix for full regression model including model fit statistics).

To compare the effects of placement and approach, a second regression analysis excluding the misinformation-only condition was performed. This analysis confirms that logic-focused approaches reduced plant myths significantly more than fact-based approaches ($b=-.42$, $S.E.=.17$, $p=.01$; RQ1a). Second, corrections that appear after the myth (debunking) produced lower plant misperceptions than corrections before the misinformation (prebunking, $b=-.47$, $S.E.=.17$, $p=.01$; RQ2a). The interaction between message features—that is, between placement and fact/logic focus—was not significant ($b=.32$, $S.E.=.24$, $p=.19$), per RQ3a.

**Misinformation credibility**

We observed a similar pattern for assessments of the credibility of the misinformation post. The logic-focused prebunking ($b=-.47$, $S.E.=.17$, $p=.01$) and the logic-focused debunking ($b=-.45$, $S.E.=.17$, $p=.01$) significantly reduced the credibility of the misinformation versus when correction was absent, with the fact-focused debunking producing a marginally significant negative relationship ($b=-.30$, $S.E.=.18$, $p=.09$). However, the fact-focused prebunking ($b=-.04$, $S.E.=.17$, $p=.82$) again produced non-significant differences in the credibility of the misinformation compared to the misinformation-only condition.

Turning to the comparison of the correction conditions (without the misinformation-only condition), the logic-focused corrections again outperform the fact-focused corrections in reducing the credibility of the misinformation post ($b=-.43$, $S.E.=.16$, $p=.01$), per RQ1b. Placement by itself has no significant effects on the credibility of the misinformation ($b=-.26$, $S.E.=.17$, $p=.12$), per RQ2b, nor does it interact with correction type to produce effects ($b=.28$, $S.E.=.23$, $p=.24$), per RQ3b.
Correction credibility

For correction credibility, we only examine the effects of the two experimental factors, as participants could not rate the credibility of the correction response in the misinformation-only condition. The logic-focused corrections receive lower credibility ratings than the fact-focused corrections (b = -.60, S.E. = .14, p < .001), per RQ1c, but this effect depends on placement (b = .51, S.E. = .21, p = .01), per RQ3d. This interaction demonstrates that the difference between the credibility of the fact-focused versus logic-focused corrections largely occurs when they appear before the misinformation. Indeed, the logic-focused prebunking is seen as less credible than all the other types of correction and the fact-focused prebunking is seen as more credible than either the logic-focused prebunking or logic-focused debunking. Placement alone does not exert a significant effect on correction credibility (b = -.23, S.E. = .15, p = .12), per RQ2c.

Mediation via misinformation credibility

Finally, we test whether the credibility of the misinformation (RQ4a) and correction (RQ4b) serve as mediators between the corrections and participants’ plant misperceptions. We again use two analyses to test this question. First, we compare the correction conditions to the misinformation conditions and examine whether the effects on plant misperceptions are channeled through evaluations of the credibility of the misinformation using Model 4 (simple mediation model) in PROCESS version 3.3 (Hayes, 2017).

We note a positive relationship between the credibility of the misinformation and plant misperceptions (b = .54, S.E. = .03, p < .001). We find that the indirect pathways between both of the logic-focused corrections (both prebunking and debunking) and plant misperceptions via misinformation credibility are significant, while these indirect pathways via either of the fact-focused corrections are not significant (see Table 1). In other words, the logic-focused
corrections work to reduce plant misperceptions by reducing the credibility of the misinformation post, but fact-focused corrections do not. Moreover, the direct effects of the correction on plant misperceptions remain significant for three of the four correction messages (excluding the fact-focused prebunking). These analyses suggest that logic-focused corrections reduce misperceptions in part by reducing the credibility of the misinformation. However, both the logic-focused corrections, as well as the fact-focused debunking, exert a direct impact on plant misperceptions that is not explained by misinformation credibility.

**Mediation via misinformation and correction credibility**

To consider the possible mediating role of the credibility of the correction in conjunction with the credibility of the misinformation, we exclude the misinformation-only condition (where participants did not see and could not rate the credibility of a correction). For these analyses, we use the fact-focused prebunking as the excluded category, as it did not produce significantly different plant misperceptions from the misinformation-only condition, using PROCESS version 3.3, model 4.

In this model, both the credibility of the misinformation \( b=.56, SE=.03, p<.001 \) and the credibility of the correction \( b=-.16, SE=.04, p<.001 \) are associated with plant misperceptions, in (expected) opposite directions (see Table 2). We again see that the logic-focused prebunking and logic-focused debunking reduce plant misperceptions via lower credibility assessments of the misinformation post as compared to the fact-focused prebunking. However, there is a significant positive relationship via correction credibility for both logic-focused corrections, as they were rated as lower in credibility than the fact-focused prebunking. In other words, the logic-focused corrections reduced plant misperceptions by reducing perceptions of the credibility of the misinformation post, but in fact increased plant misperceptions via the lower credibility of
these correction posts versus the fact-focused prebunking. Finally, there remains a significant
*direct* effect on plant misperceptions (outside of credibility perceptions), by which exposure to
logic-focused prebunking, logic-focused debunking, and fact-focused debunking result in lower
plant misperceptions compared to a fact-focused prebunking. Taking all three pathways—direct,
indirect via misinformation post credibility, and indirect via correction post credibility—into
account, logic-focused prebunking, logic-focused debunking, and fact-focused debunking all
produce lower plant misperceptions compared to the fact-focused prebunking.

**Discussion**

This study addresses several important gaps in the literature. First, we examine
misinformation and correction of a controversial political topic (climate change) on a popular but
understudied visual platform—Instagram—using a humorous approach that may be especially
well suited to visual social media platforms. Second, we compare two features of correction:
whether it takes a logic-focused or fact-focused approach and whether the correction occurs
before (prebunking) or after (debunking) exposure to the misinformation. Third, we test whether
message credibility of both the misinformation and the correction can mediate the effectiveness
of these strategies in shaping issue attitudes. Our results suggest that observational correction is
effective on visual platforms like Instagram, even when the correction is not directly threaded to
the misinformation post, but both the correction type and its placement matter for these effects.

Overall, we find that logic-focused corrections, which highlight the rhetorical flaws of the
misinformation claim, outperform fact-focused corrections, which debunk the claim itself by
highlighting scientific facts on the issue, because the logic-focused corrections are impervious to
placement. While the logic-focused and fact-focused corrections are equally effective when seen
*after* the misinformation post, only the logic-focused corrections reduce misperceptions when
they appear before the misinformation, offering more flexibility in their application. This is particularly important because we cannot know whether a user has been exposed to misinformation when sharing corrective information on social media. Knowing that a strategy will work whether it is preemptive or reactive is therefore very useful.

Our mediation analyses provide insight into the process by which logic-focused corrections function differently than fact-focused corrections. Logic-focused corrections produce an indirect effect on plant misperceptions through reduced credibility ratings for the misinformation post—an indirect pathway that is not significant for those viewing a fact-focused condition (although it is directionally consistent). However, this indirect pathway for logic-focused corrections via the credibility of the misinformative post is only part of the story—a significant direct pathway remains, reducing misperceptions. More research is needed into what mechanisms explain this direct pathway in reducing misperceptions for both the logic-focused and fact-focused approaches, but this study suggests that perceptions of message credibility offer promise as a mechanism to explain some of the effectiveness of logic-focused corrections.

Rhetorical strategies offer two benefits for correcting misinformation. First, they function as a proactive strategy; one does not need to wait until they see misinformation being spread to correct it. Second, scholars have suggested that rhetorical approaches generalize across issues (Cook et al. 2017; Schmid and Betsch 2019). Once an individual is familiar with common misinformation techniques—for example, oversimplification or use of fake experts (Cook et al. 2018; van der Linden et al., 2017)—they can apply this knowledge to other misinformation that uses the same technique. While we did not test whether the logic-focused correction is transferrable in this study, the fact that it functions by reducing the credibility of misinformation message supports this argument.
Correction placement and type on Instagram

But while the logic-focused corrections may appear superior to the fact-focused corrections given their flexibility as a prebunking or debunking strategy, fact-focused corrections have merit. The fact-focused corrections reduced plant misperceptions when they occurred after the misinformation post, consistent with existing research into observational correction on social media that previously relied on fact-focused debunking (Bode and Vraga 2015; Smith & Seitz, 2019; Vraga and Bode 2018). Additionally, the fact-focused correction offers a second benefit: it was rated higher in credibility than the logic-focused correction, especially when seen before the misinformation. One explanation may be that the fact-focused prebunking was judged before participants viewed the misinformation and sharing of facts is widely valued. In contrast, the logic-focused prebunking may have seemed out of place when appearing before misinformation. For individuals or organizations sensitive to concerns about their credibility in responding to misinformation, a fact-focused approach may hold greater appeal. Moreover, fact-focused corrections may be more suitable for content experts (Schmid and Betsch, 2019), broadening the number and types of people who can effectively respond to misinformation.

Additionally, while this study examined fact-focused and logic-focused corrections as separate factors, future research could explore whether the combination of fact-focused and logic-focused corrections before and after misinformation best reduces misperceptions. However, Schmid & Betsch (2019) found no evidence of complex combinations being more effective than single strategies.

This study builds on previous work into observational correction in three ways. First, it extends the platforms on which observational correction operates to Instagram: one of the most popular social media platforms in the United States (Perrin and Anderson 2019) and a platform that prioritizes visual content. Efforts to encourage observational correction on Instagram seem
warranted. Second, it demonstrates that corrections need not directly reply to misinformation, so long as they appear immediately afterwards (or before, when using a logic-focused approach). It is impossible to know exactly when or where people have seen misinformation; the effectiveness of a corrective post likely depends on its relative proximity. Future research should test how close the two posts must be for corrections to work. Third, despite previous research suggesting that fact-focused observational correction from users requires at least two sources with links to expert evidence (Vraga and Bode 2017; 2018), this study suggests this requirement is not absolute. It may be the relative obscurity of the plant misinformation being studied—as compared to the prominence of the Zika virus in previous studies—contributes to its effectiveness. Alternatively, it may be the expertise signal of the user offering the correction (described as Prof.Mason.Green although no other credibility cues are provided) facilitates correction, as expert organizations only need a single correction to reduce misperceptions on social media (Vraga and Bode 2017). Simply having “Prof” in the title may confer expertise, although future research should investigate the boundaries of expertise, given work on the importance of source cues (Bolsen et al. 2019b). Of course, fact-focused correction offered in this study may have been strengthened if it included a link to an expert source—potentially negating the failure of the fact-focused prebunking to reduce misperceptions.

This study has several limitations. First, we tested a single exposure to an unknown source of both the misinformation and the correction on an artificial feed. We do not know how long the effects of the correction will be maintained, or how these effects would differ depending on the source of the misinformation and the correction or people’s native use patterns. While our study suggests that these results hold up well on platforms (like Instagram) where people often encounter posts from people they may not personally know (Phua et al. 2017), more research is
needed into source cues. Second, we tested corrections that used humorous cartoons as part of their correction strategy (and participants found all corrections equally humorous—about a 4.5 on a scale of 1 to 7). While this choice was deliberate to match the norms of the platform, we did not compare these effects to a non-humorous approach. Previous research has found non-humorous corrections were equally or more effective in reducing misperceptions than humorous ones (Authors 2019; Young et al. 2018), although testing humor in corrections has been rare and represents an important area for future research. Third, our logic-focused correction did contain some factual information, which may have strengthened its effects. While these facts are included to highlight the rhetorical flaws as done in past research using logic-focused corrections (e.g., Schmid & Betsch, 2019), future research should isolate logic-focused corrections, absent any factual information. Finally, while the overall issue of climate change is quite political, the specific myth about carbon dioxide we tested is not politicized\(^3\), which may have made it more likely to find effects (Bolsen and Druckman 2018).

Ultimately, this study offers insight for those interested in how to best correct misinformation on Instagram or other visual social media platforms. Our recommendations are to use logic-focused humor corrections, which are flexible enough to reduce misperceptions whether they come before or after exposure to misinformation. Fact-focused corrections also have merit, especially when responding to misinformation already circulating on the platform. Users should be encouraged to respond and share corrective messages on visual platforms to ensure that high quality information dominates the online landscape.

\(^3\) Supplemental analysis shows no differences in how Republicans and Democrats reacted to the stimuli.
References

Authors (2019).

Authors (n.d.).


Bolsen, Toby, and James N. Druckman. 2018. "Do partisanship and politicization undermine the
Correction placement and type on Instagram

impact of a scientific consensus message about climate change?" Group Processes &
Intergroup Relations 21: 389-402.

Bolsen, Toby, Risa Palm, and Justin T. Kingsland. 2019a. "Counteracting Climate Science
Politicization With Effective Frames and Imagery." Science Communication 41: 147-
171.

Bolsen, Toby, Risa Palm, and Justin T. Kingsland. 2019b. "The Impact of Message Source on
the Effectiveness of Communications About Climate Change." Science Communication
41: 464-487.

Boussalis, Constantine and Travis G. Coan. 2016. “Text-mining the Signals of Climate Change
Doubt.” Global Environmental Change 36: 89-100.

Brewer, Paul R and Jessica McKnight. 2015. “Climate as Comedy: The Effects of Satirical
Television News on Climate Change Perceptions.” Science Communication 37: 635-657.

Castillo, Carlos, Marcelo Mendoza, and Barbara Poblete. 2013. “Predicting Information

Clayton, Katherine, Spencer Blair, Jonathan A. Busam, Samuel Forstner, John Glance, Guy
Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False
Stories on Social Media." Political Behavior: 1-23.

Misinformation Through Inoculation: Exposing Misleading Argumentation Techniques

Cook, John, Naomi Oreskes, Peter T. Doran, William RL Anderegg, Bart Verheggen, Edward


Correction placement and type on Instagram

May Influence Credibility Assessment and Belief Change." Journalism & Mass Communication Quarterly 73: 974-991.


Tables and Figures

Table 1: Indirect pathways of corrections versus misinformation-only conditions in predicting plant misperceptions via misinformation credibility (n = 1005)

<table>
<thead>
<tr>
<th></th>
<th>Plant Misperceptions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
</tr>
<tr>
<td>Indirect via misinformation credibility</td>
<td></td>
</tr>
<tr>
<td>Logic-focused prebunking</td>
<td>-.25</td>
</tr>
<tr>
<td>Fact-focused prebunking</td>
<td>-.02</td>
</tr>
<tr>
<td>Logic-focused debunking</td>
<td>-.24</td>
</tr>
<tr>
<td>Fact-focused debunking</td>
<td>-.16</td>
</tr>
<tr>
<td>Direct pathway</td>
<td></td>
</tr>
<tr>
<td>Logic-focused prebunking</td>
<td>-.31</td>
</tr>
<tr>
<td>Fact-focused prebunking</td>
<td>-.13</td>
</tr>
<tr>
<td>Logic-focused debunking</td>
<td>-.46</td>
</tr>
<tr>
<td>Fact-focused debunking</td>
<td>-.45</td>
</tr>
<tr>
<td>Total effects</td>
<td></td>
</tr>
<tr>
<td>Logic-focused prebunking</td>
<td>-.56</td>
</tr>
<tr>
<td>Fact-focused prebunking</td>
<td>-.15</td>
</tr>
<tr>
<td>Logic-focused debunking</td>
<td>-.71</td>
</tr>
<tr>
<td>Fact-focused debunking</td>
<td>-.61</td>
</tr>
</tbody>
</table>

Notes. Unstandardized beta coefficients reported; significant effects bolded when 95% confidence interval does not include zero, $p<.05$
Table 2: Indirect pathways of corrections versus fact-focused prebunking conditions in predicting plant misperceptions via misinformation and correction credibility (n = 821)

<table>
<thead>
<tr>
<th></th>
<th>Plant Misperceptions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
</tr>
<tr>
<td>Indirect via misinformation credibility</td>
<td></td>
</tr>
<tr>
<td>Logic-focused prebunking</td>
<td>-.24</td>
</tr>
<tr>
<td>Logic-focused debunking</td>
<td>-.23</td>
</tr>
<tr>
<td>Fact-focused debunking</td>
<td>-.15</td>
</tr>
<tr>
<td>Indirect via correction credibility</td>
<td></td>
</tr>
<tr>
<td>Logic-focused prebunking</td>
<td>.10</td>
</tr>
<tr>
<td>Logic-focused debunking</td>
<td>.05</td>
</tr>
<tr>
<td>Fact-focused debunking</td>
<td>.04</td>
</tr>
<tr>
<td>Direct effects</td>
<td></td>
</tr>
<tr>
<td>Logic-focused prebunking</td>
<td>-.27</td>
</tr>
<tr>
<td>Logic-focused debunking</td>
<td>-.38</td>
</tr>
<tr>
<td>Fact-focused debunking</td>
<td>-.36</td>
</tr>
<tr>
<td>Total effects model</td>
<td></td>
</tr>
<tr>
<td>Logic-focused prebunking</td>
<td>-.41</td>
</tr>
<tr>
<td>Logic-focused debunking</td>
<td>-.56</td>
</tr>
<tr>
<td>Fact-focused debunking</td>
<td>-.46</td>
</tr>
</tbody>
</table>

Notes. Unstandardized beta coefficients reported; significant effects bolded when 95% confidence interval does not include zero, $p<.05$
Figure 1: Effects of correction type and placement on plant misperceptions

![Bar chart showing effects of correction type and placement on plant misperceptions.](image-url)
Figure 2: Effects of correction type and placement on misinformation and correction credibility
Appendix: Experimental Images

<table>
<thead>
<tr>
<th>Misinformation post</th>
<th>Fact-focused correction</th>
<th>Logic-focused correction</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Misinformation post" /></td>
<td><img src="image2" alt="Fact-focused correction" /></td>
<td><img src="image3" alt="Logic-focused correction" /></td>
</tr>
</tbody>
</table>

**Note.** The order of the correction posts (before vs. after) were changed for the placement manipulation.