Cite as: Kim, S. C., Vraga, E. K., & Cook, J. (2020). An Eye Tracking Approach to Understanding Misinformation and Correction Strategies on Social Media: The Mediating Role of Attention and Credibility to Reduce HPV Vaccine Misperceptions. *Health Communication*.

An eye tracking approach to understanding misinformation and correction strategies on social media: The mediating role of attention and credibility to reduce HPV vaccine misperceptions

Abstract

This study uses an unobtrusive eye tracking approach to examine understudied psychological mechanisms—message attention and credibility—when people are exposed to misinformation and correction on social media. We contrast humor versus non-humor correction strategies that point out rhetorical flaws in misinformation regarding the HPV vaccine, which was selected for its relevance and impact on public health. We randomly assigned participants to one of two experimental conditions: humor correction versus non-humor correction. Our analyses revealed that the humor correction increased attention to the image portion of the correction tweet, and this attention indirectly lowered HPV misperceptions by reducing the credibility of the misinformation tweet. The study also found that the non-humor correction outperformed the humor correction in reducing misperceptions via its higher credibility ratings. Practical implications for correcting misinformation on social media are discussed.

Keywords: Social media; Misinformation; Observational correction strategies; Humor appeals; Attention; Message credibility; HPV vaccine misperceptions; Eye tracking

An eye tracking approach to understanding misinformation and correction strategies on social media: The mediating role of attention and credibility to reduce HPV vaccine misperceptions

Introduction

Amid concerns about misinformation on social media for a variety of health, political, and science topics, special attention has been paid to vaccine misinformation. Studies have documented the prevalence and popularity of misinformation surrounding vaccination across social media platforms (e.g., Brionanowski et al., 2018; Guidry et al., 2015). Strategies to *correct* vaccine misinformation have also received substantial attention. Correcting misinformation once it is firmly established can be quite difficult (Bode & Vraga, 2015; Lewandowsky et al., 2012; Nyhan & Reifler, 2010), as is the case for many vaccination issues, including the HPV vaccine.

On social media, the practice of *observational correction*—which occurs when seeing someone else being corrected on social media leads individuals to adopt more accurate attitudes on the issue (Vraga & Bode, 2017)—has been successful across a range of issues. Existing tests of observational correction have largely focused on factual rebuttals of the misinformation relying on expert sources (Bode & Vraga, 2015; van der Meer & Jin, 2019; Vraga & Bode, 2017), but an alternative strategy of logic-based correction has been the focus of renewed theorizing (e.g., Vraga et al., 2019; Cook et al., 2017; Schmid & Betsch, 2019).

This study tests the effectiveness of a logical correction for observational correction on social media by contrasting two techniques: a humorous correction that uses parallel argumentation to underscore the logical fallacy inherent in the misinformation versus a non-humorous argument deconstruction naming the specific logical fallacy. Humorous messages are an engaging format for scientific information especially among audiences who are disengaged from an issue (Becker & Waisman, 2013; Brewer & McKnight, 2017; Skurka et al., 2018) and

can boost message credibility (Vraga et al., 2014). Moreover, visual imagery and humor can encourage greater message attention and virality on social media (Campo et al., 2013, Vraga et al., 2016), which may help these corrections garner attention.

Despite the interest in strategies for correcting misinformation on social media, however, less emphasis has been placed on the *mechanisms* by which corrections function. Based on the past literature, two mechanisms stand out: (1) attention to the misinformation itself and the correction, and/or through (2) the credibility of the misinformation and its correction. To fill the gap in the existing literature, this study uses an unobtrusive eye tracking approach to precisely measure how much *attention* audiences pay to a misinformation and a correction message, and how attention is shaped by the correction strategy employed (humor versus non-humor). Then, we explore *how* attention and credibility perceptions mediate the effectiveness of these correction strategies in reducing misperceptions on the timely topic of HPV vaccination.

Misinformation and Different Types of Corrections

Misinformation, which we define as information that counters expert evidence and knowledge on an issue (Bode & Vraga, 2015; Lewandowsky et al., 2012; Nyhan & Reifler, 2010), can lead people to make suboptimal health decisions (e.g., Jolley & Douglas, 2014). Many scholars have highlighted the difficulty in correcting misperceptions, especially when beliefs are entrenched (Lewandowsky et al., 2012; Nyhan & Reifler, 2010; Thorson, 2016). Even when people accept corrected information, the effects of misinformation on attitudes can persist, creating "belief echoes" or a "continued influence effect" (Ecker et al., 2010; Thorson, 2016).

On social media, observational correction techniques show particular promise. That is, when users correct other users sharing misinformation on social media, it may not only affect the

person who was corrected (Margolin et al., 2018), but also reduce misperceptions among the community seeing the correction (Vraga & Bode, 2018).

Three main types of corrections have been explored in existing research: fact-based, source-based, and logic-based corrections. Fact-based corrections counter misinformation by explaining relevant facts that conflict with misleading claims. Source-based corrections highlight the motives or disreputable traits of misinformation sources. Most correctional research has focused on fact-based corrections with links to expert sources (Vraga & Bode, 2018) or through experts themselves (van der meer & Jin, 2019; Vraga & Bode, 2017).
In contrast, logic-based corrections use rhetorical approaches to highlight the logical flaws in misinformation (Cook et al., 2017; Schmid & Betsch, 2019). One advantage of logic-based refutation is their generalizability: an explanation of a misleading fallacy can apply to a range of topics while a fact-based refutation is only relevant to a specific topic (Schmid & Betsch, 2019). This is consistent with research finding that a single inoculating message can convey an "umbrella of protection" across multiple misleading arguments (Parker et al., 2012). Logic-based corrections have been found to be just as or more effective as fact-based refutations in countering misinformation (Cook, et al., 2017; Parker et al., 2012; Schmid & Betsch, 2019; Vraga et al., 2020).

In this study, we test two ways to advance a logic-based correction, based on Cook, Ellerton, & Kinkead (2018)'s review of misinforming arguments regarding climate change. One approach involves deconstructing misleading statements into their constituent argument structure of premises and conclusion to identify any potential reasoning fallacies in the argument. A second approach is parallel argumentation, where the logical structure of a misinforming argument is applied to an extreme situation in order to clearly explain the flawed logic (Juthe,

2009). The latter approach is conducive to humorous treatments, opening opportunities to take advantage of the benefits of humorous science communication.

Humor Appeals as a Type of Correction

A number of studies have explored the potential benefits of humor in health and science communication with most studies applying humor in a non-correctional context. Bore and Reid (2014) described humor in science messages as "sugaring the pill" (p. 470), employing comedic devices to facilitate serious discourse. Humor is especially likely to engage audiences who are less engaged on an issue (Becker & Waisman, 2013; Brewer & McKnight, 2017). Humorous communication enhances the perceived credibility of the messenger (Vraga et al., 2014), as well as reduces counter-arguing, possibly because audiences who spend more cognitive effort on "getting the joke" have less cognitive resources available for argument scrutiny (Young, 2008). Moyer-Gusé, Robinson, and McKnight (2018) found that a humorous message about vaccination reduced reactance and increased openness to information about the severity of the measles virus, although Young et al. (2018) found that a non-humorous fact-checking video was equally effective in reducing misperceptions as a humorous video.

Humor in science communication may also lead to negative effects. While humor can make difficult subjects more accessible, this can also result in people taking the issue less seriously (McGraw et al., 2015; Nabi et al., 2007). Skurka and his colleagues (2018) found tradeoffs in a humorous message about climate change, by indirectly increasing climate-related behavior intent through perceptions of humor, but indirectly decreasing behavior intent through reduced fear and perceived message informativeness.

Research into humorous inoculating treatments have found mixed results, with audience segments responding differently to the same humorous messages (Compton, 2018), but most past

scholarly work on humorous science communication does not test responses to misinformation (but see Young et al., 2018). When humor is added to a correction message, it may change how people process the information, such as attention patterns, credibility assessments, and ultimately, their attitudes regarding the topic.

Studying Attention on Social Media

While existing research has proposed and tested a number of different correction strategies, the question of the *mechanisms* by which corrections reduce the influence of misinformation has been understudied. One mechanism that is potentially significant is *attention*, especially online where diverse content competes to engage audiences (Stroud, 2017).

While attention to content has been broadly explored with social media content (e.g., Sulflow et al., 2019; Vraga et al., 2016), it has not been applied to misinformation and correction on social media. Traditional studies of correction measure whether *exposure* to a correction message predicts change in attitudes, without testing the *degree of attention* audiences paid to that message. This is a critical oversight, as attention is a precursor to a number of processes necessary to successful correction; we would not expect a message that people do not attend to will impact their attitudes.

Therefore, a first step is to examine whether different features of a correction—specifically a non-humorous argument deconstruction of reasoning fallacies versus a humorous refutation in the form of a parallel argument—affect processing. Our previous research found that the non-humorous explanation reduced HPV misperceptions more than the humorous refutation, largely among those who originally held stronger misperceptions on the issue (Vraga et al., 2019). We build on this research with a new data collection using eye tracking to test whether these two approaches to correct misinformation lead to different attention patterns.

RQ1: Will a humorous correction strategy lead to greater attention to a) text and b) image portions of the misinformation tweet than a non-humorous correction strategy?

RQ2: Will a humorous correction strategy lead to greater attention to a) text and b) image portions of the correction tweet than a non-humorous correction strategy?

Attention is often a fundamental—but understudied—component of models of information processing and persuasion. For example, the Elaboration Likelihood Model (ELM) distinguishes between a peripheral, low-effort route by which persuasion can occur, in opposition to a central, effortful route that involves "attend[ing] to the appeal" (Petty & Cacioppo, 1986, p. 128) before affecting audience attitudes. Likewise, the Limited Capacity Model of Mediated Message Processing (LC4MP) suggests that processing media messages involves three subprocesses: encoding, storage, and retrieval, with encoding deeply reliant on attention processes (Lang, 2000; 2006). These models place attention as a precursor to learning (e.g., storage and retrieval in the LC4MP) and persuasion (e.g., the ELM)—critical components of an effective corrective response.

While degree of attention has not been explicitly measured in correction research to our knowledge, studies on selective exposure (which inherently drives attention) to correction reinforce the importance of this variable. People are more likely to click on and share fact-check articles that agree with their beliefs (Hameleers & van der Meer, 2019; Shin & Thorson, 2017). Additionally, incongruent fact-checks (e.g., that counter one's political beliefs) have larger effects on attitudes among those who choose to read the fact-check than those who avoid it (Hameleers & van der Meer, 2019), suggesting that some degree of attention is required.

Therefore, existing research hints at the important role attention may play in shaping the effectiveness of corrections, but does not explicitly measure attention as a mediating role

between exposure to a correction message and a reduction of misperceptions as a response. Building from these theories of persuasion leads to two possible models. First, the LC4MP suggests message attention to both the misinformation post and the correction may *directly* mediate the effects of correction on updated HPV attitudes by shaping whether people encode and later retrieve the (inaccurate and accurate) information presented via the posts. If people attend to the misinformation message, it should directly affect their ability to store and retrieve its information, which we suggest is an important precursor to attitude change.

Second, we propose that message attention may have an indirect effect on misperceptions via credibility assessments of the two messages. Specifically, the ELM suggests that more effortful processing (of which high attention may be one marker) should lead individuals to rely more on message quality when interpreting the information. As such, high attention to posts may lead people to rate the misinformation message as less credible given its low argument quality and the correction tweet as more credible.

H1: Higher attention to the misinformation tweet will lead to a) lower credibility ratings of the misinformation tweet, b) higher credibility ratings of the correction tweet, and c) lower HPV misperceptions.

H2: Higher attention to the correction tweet will lead to a) lower credibility ratings of the misinformation tweet, b) higher credibility ratings of the correction tweet, and c) lower HPV misperceptions.

Misinformation and Correction Credibility

Many theoretical traditions recognize credibility as playing an important orienting role towards messages, including boosting its persuasive power (Petty & Brinol, 2008; Petty & Cacioppo, 1986; Sulflow et al., 2019). As noted above, the ELM highlights the role of both

attention and credibility in determining whether a message is successful in shifting audience attitudes towards the target.

Existing misinformation studies have largely focused on the credibility of the *source* of the correction. A highly credible source is largely assumed to be an important component of successful debunking strategies (Lewandowsky et al., 2012), and many studies have relied on expert sources for correction responses, including journalists (Nyhan & Reifler, 2010; Thorson, 2016), expert organizations (Vraga & Bode, 2017), and fact-checking organizations (Hameleers & van der Meer, 2019). Indeed, in direct comparisons, research has found that expert sources like the CDC are more effective than other users in reducing misperceptions and anxiety, which has been theoretically attributed to their credibility in delivering the message (van der Meer & Jin, 2019; Vraga & Bode, 2017), echoing a meta-analysis that finds high credibility sources are more persuasive (Pornpitakpan, 2004).

The credibility literature distinguishes between multiple types of credibility—the credibility of the *source* (as described above) versus the credibility of the *message* itself (Metzger et al., 2003). Message credibility is largely considered in terms of the content features of the message, often measured by ratings of the credibility, accuracy, and trustworthiness of the message (e.g., Flanagin & Metzger, 2000). This also aligns with the concept of *argument strength* of the message from the ELM, which operates via the central route (Petty & Brinol, 2008; Petty & Cacioppo, 1986).

While there is overlap between source and message credibility (see Metzger et al., 2003 for a discussion), they are conceptually distinct. We focus on message credibility in this study, as we expect that our correction strategies may be effective by shifting perceptions of the credibility of the misinformation message by highlighting its logical flaws while holding source

credibility constant across the messages. Given limited research comparing humorous logical refutations to non-humorous argument deconstruction, we ask:

RQ3: Will a humorous correction strategy lead to a) lower credibility ratings of the misinformation tweet and b) higher credibility ratings of the correction tweet than a non-humorous correction strategy?

RQ4: Will a humorous correction strategy lead to lower HPV misperceptions than a non-humorous correction strategy?

While credibility is important on its face, we also explore its potential to serve as a mediator to reduce misperceptions. Vraga and Bode (2018) find that higher perceptions of the correction message credibility reduce misperceptions regarding the causes of the Zika virus on Twitter, while Guillory and Geraci (2013) find source trustworthiness is more important than expertise in debunking political misinformation. However, both of these studies focus on the credibility of the correction message, rather than comparing it to the misinformation message being targeted. Responding to this gap in the literature, we examine the credibility of both the misinformation and the correction posts. We expect credibility perceptions of these posts to work in opposition: lower credibility of the misinformation tweet and higher credibility of the correction responses should both reduce misperceptions on HPV vaccine.

H3: Lower credibility ratings of the misinformation tweet will lead to lower HPV vaccine misperceptions.

H4: Higher credibility ratings of the correction tweet will lead to lower HPV vaccine misperceptions.

Finally, we expect that attention precedes evaluations of both misinformation and correction tweet credibility. Individuals must process or pay attention to information to form an

opinion regarding message credibility. Past literature finds that message credibility influences misperceptions in different topical contexts (e.g., Guillory & Geraci, 2013; Vraga & Bode, 2018). Moreover, graphical refutations (Nyhan & Reifler, 2019) and visual imagery (Bolsen et al., 2019) have been shown to be particularly effective in countering misinformation on scientific and political topics. This may explain why images that support information in text have generated more attention, more processing, greater comprehension, and stronger effects toward behavior changes than textual messages in health communication (e.g., Houts et al., 2006; Lazard & Atkinson, 2015). Given limited research evidence in this area, we propose the following research question:

> **RQ5:** Will there be mediating effects of message attention and credibility when investigating the relationship between logic-based correction strategies and the HPV misperceptions, and how this mediating process will differ for text and image portions of the tweets?

Refer to Figure 1 for visualization of the conceptual model.

[FIGURE 1 ABOUT HERE]

Methods

Experimental Design

The study employed a between-subjects experiment, where participants were randomly assigned to one of the two experimental conditions: humor versus non-humor correction[1]. The

[1] The experiment also included misinformation only condition as control (n = 31), but for the purpose of this study, these participants were excluded because they did not receive any corrections.

two conditions adopted different communication approaches, with each approach specifically suited to either the humor or non-humor condition. The non-humor condition visually deconstructed the misinforming argument, highlighting the logical fallacy of "correlation implies causation." The humor condition adopted a parallel argumentation approach, where the "correlation implies causation" fallacy was applied to an alternative, transparently false situation. These two approaches were chosen to maximize the difference in humor levels across the two conditions as well as the realism of the stimuli.

Twitter was chosen as the social media platform for the message manipulation due to its interactive and open nature. In both correction conditions, an original tweet contained misinformation about HPV vaccine, claiming that it caused auto-immune symptoms like paralysis with an accompanying image. The correction was offered as a reply tweet, with text explaining the logical fallacy in the misinformation in the original tweet. The text was identical in both correction versions. We manipulated the image that appeared after this text in the correction tweets to reinforce the explanation of the logical flaw using two competing strategies. Specifically, in the humor correction, we featured a cartoon using parallel argumentation, whereas the non-humor correction condition used an infographic that visualized the misinformation argument structure and named the specific logical fallacy. The same gender-neutral names and generic user icons were used across all conditions to avoid potential influence of gender bias or source credibility.

Participants

Study participants were students from a major mid-Atlantic university in the United States over the age of 18, who could speak and understand English. In total, 92 participants completed the study but only participants (n = 61) who viewed both the misinformation and the

correction were included in the analyses[2]. Participants were 37.7 % male, 44.3 % White, and the mean age was 20 years old (*SD* = 1.87, *Min* = 18, *Max* = 25).

Eye Tracking Setup and Procedures

The study was conducted in three stages: All participants in the study completed a pretest survey online before coming to an eye tracking research lab. At the completion of the pretest survey, participants were directed to make an appointment for the lab within a week time frame.

Once in the lab, each participant was randomly assigned to one of the two experimental conditions and viewed ten social media messages, while their eyes were tracked with Tobii Pro X3-120 screen-based portable eye tracker (Tobii Technology, Stockholm, Sweden). This eye tracker captures eye movements with a sampling rate of 120 Hz, and was attached at the bottom of a 17-inch monitor (with the plastic edge) on a desk with stationary chair. The screen resolution was 1,440 x 900 pixel. Unlike other head-mounted eye trackers, participants could move their heads and body freely: Freeing participants from sitting completely still to track their eyes reduces the artificiality of a lab setting and improves external validity of the study.

Calibration was completed before exposure to message stimuli and required participants to follow a black dot inside of a larger red circle. Participants went through the partial or full calibration up to three times, and if a participant reached a third unsuccessful calibration ($n = 6$), he or she was excused from the study. The eye movement was classified as a fixation by the software if the eyes moved less than 30 degrees per second (Olsen, 2012).

During the eye tracking experiment, participants were encouraged to review each tweet

[2] Our original sample size of N = 92 is substantially higher than the average of 74 participants in communication-relevant eye tracking studies (King et al., 2019).

for at least ten seconds, but there were no restrictions on how much time spent for each message, allowing message viewing at their own pace. Participants were exposed to one tweet (i.e., the manipulation) about the HPV vaccine along with nine dummy tweets to mask the specific study purpose. Dummy tweets imitated tweets participants would see on Twitter, including ordinary activities like reading a science fiction, making chocolate chip cookies, and hiking with comfortable shoes, along with relevant images. The images included in the original misinformation tweet were adapted from a media article, and infographic and cartoon images in the correction reply tweets were created and further refined by the third author. Tweet order was randomized within participants to control for potential order effects.

After reviewing the manipulated tweets, participants completed a posttest survey through Qualtrics, answering questions about HPV attitudes, misperceptions, and message credibility. At the study completion, participants were compensated with either course credit or $5. One randomly selected participant was awarded a $20 cash prize in addition to their chosen option.

Eye Tracking Measures

With eye tracking technology, attention was operationalized and measured as *total fixation duration* (TFD) per defined Areas of Interest (AOIs). TFD refers to the sum of all fixation durations within a given AOI (Just & Carpenter, 1976). An AOI is a pre-defined area in a message a researcher is interested in examining based on his or her research questions (Holmqvist et al., 2011). For this study, text and image areas of both misinformation and correction tweets across all conditions were defined as AOIs and participants' viewing patterns of these AOIs were captured and processed by the eye tracking software. See Figure 2 for further details on the AOIs.

[FIGURE 2 ABOUT HERE]

Attention to misinformation tweet. Total fixation duration (TFD), which is the time spent looking at the original misinformation tweet, was used to assess message attention. TFD to the text (M = 6.53, SD = 3.49) and image (M = 2.79, SD = 2.26) areas of the misinformation tweet was examined separately in milliseconds, which we report in seconds for readability.

Attention to correction tweet. TFD was also used to measure attention to the text (M = 8.43, SD = 4.45) and image (M = 5.57, SD = 3.69) areas of the correction tweet.

Other Measures

Misinformation tweet credibility. To measure credibility, participants answered on a 7-point scale: "How would you describe the content of the original tweet you've just reviewed?" Credible, informative, and accurate were provided as semantic differential scales (adapted from Vraga & Bode, 2018). A credibility index was created by averaging these three items, with a higher score meaning greater credibility (Cronbach's α = .88, M = 3.10, SD = 1.42).

Correction tweet credibility. The same credibility questions used for misinformation tweet credibility were used to measure the "content of the reply tweet" on a 7-point scale (Cronbach's α = .80, M = 4.51, SD = 1.24).

HPV vaccine misperceptions. On a 7-point scale from Strongly Disagree (1) to Strongly Agree (7), this construct was measured with the following six statements, "The HPV vaccine protects against cervical cancer (reverse coded)," "I worry the HPV vaccine might negatively affect my body," "The HPV vaccine is safe for the public (reverse coded)," "The HPV vaccine leads to auto-immune symptoms," "The HPV vaccine can protect children from getting a fatal disease (reverse coded)," and "The HPV vaccine may result in long-term side effects that are not known yet." A misperception index was created by averaging these statements, with a higher score meaning greater misperceptions about the HPV vaccine. The pretest score was controlled

for all subsequent analyses to avoid potential confounding effects (*pretest:* Cronbach's α = .83, M = 3.31, SD = .93; *posttest:* Cronbach's α = .82, M = 3.22, SD = .99).

See Table 1 for zero-order correlations of all variables of interests.

[TABLE 1 ABOUT HERE]

Analytical Strategy

The eye tracking data were processed and exported to SPSS using the Tobii Pro Studio software (Tobii Pro, 2019). Using SPSS version 26, a series of ANCOVA and PROCESS analyses with model 80 (a serial mediation) were performed to systematically examine the proposed conceptual model (Hayes, 2017; see Figure 1). For the PROCESS results, the number of samples for bootstrap confidence intervals were 5,000 and we applied a heteroscedasticity consistent estimator for standard errors (HC3), in line with best practices (Hayes, 2017).

Results

Manipulation Check

We tested if our humor correction manipulation was, in fact, perceived as humorous to participants. After viewing the manipulated tweets, participants were asked to rate whether the correction tweet was humorous/not humorous with 7-point semantic differential scales (M = 2.92, SD = 1.90). An ANCOVA was performed, controlling for the pretest score of HPV misperceptions. There was a significant main effect of correction type on humor, $F(1, 58)$ = 13.71, p = .001, *partial* η^2= .19: People who were exposed to the humor correction (M = 3.70, SE = .31) rated the tweet more humorous than people who viewed the non-humor correction (M = 2.06, SE = .32).

Direct Effects (RQ1 through RQ4; H1 through H4)

We began by examining the effects of the correction strategy (i.e., the manipulation) on

attention to four areas: (1) the misinformation text, (2) the misinformation image, (3) the correction text, and (4) the correction image, controlling for pretest HPV misperceptions.

We find that the humorous correction produced significantly more attention to the misinformation text but not to the misinformation image. In contrast, participants paid more significantly more attention to the correction image in the humor correction as compared to the non-humor correction (see Table 2 and 3), but there were no significant differences in attention to the correction text.

Moreover, we observed no direct effects of the correction manipulation on credibility assessments of the misinformation tweet or HPV misperceptions (see Table 2). As for the correction tweet credibility, we found a significant direct effect of correction strategy on the credibility of the correction tweet, with the non-humor correction receiving higher credibility ratings (see Table 3).

Additionally, a significant negative relationship between attention to the correction image and misinformation credibility was found (H2a partially supported; see Table 2). A significant positive relationship between misinformation credibility and HPV misperceptions (H3 supported; see Table 2), and a strongly significant negative relationship between correction credibility and HPV misperceptions emerged from the analyses (H4 supported; see Table 3). Other proposed relationships between attention and credibility and HPV misperceptions were not significant (i.e., H1a through H1c, H2b, and H2c were not supported).

Indirect Effects (RQ5)

We were also interested in potential *mediating* effects of attention and message credibility on the relationship between correction strategy and HPV misperceptions. We ran two separate PROCESS mediations to test these relationships, one for each credibility as a second-

level mediator. We observed that the humor correction garnered more attention to the correction image, and this increased attention contributed to reducing the credibility of the *misinformation* tweet, and the reduced misinformation credibility ratings produced lower HPV misperceptions, producing a significant indirect effect (see Table 2). For the correction credibility, we did not find a significant mediated pathway through attention to the misinformation tweet or the correction tweet. Instead, there was a significant indirect pathway from correction strategy to HPV misperceptions, revealing those in the non-humor correction to have lower HPV misperceptions through the higher credibility ratings of the *correction* tweet than those in the humor correction (see Table 3). Other suggested indirect paths in the conceptual model were not significant. See Table 2 and 3 for detailed results.

[TABLE 2 AND 3 ABOUT HERE]

Discussion

This study used an unobtrusive eye tracking approach to systematically examine understudied psychological mechanisms people engage when they process misinformation and correction on social media. We selected a socially-contested health topic—HPV vaccination—given the salience and prevalence of misinformation on vaccination (e.g., Brionanowski et al., 2018; Guidry et al., 2015) and its importance to public health (e.g., Kim et al., 2018). We examined the effectiveness of humorous versus non-humorous correction strategies in garnering attention and the implications of such attention for perceptions of message credibility and issue misperceptions. The direct and indirect effects our study found are not only consistent with but also expand the current literature on attention, credibility, and misinformation correction (Cook et al., 2017; Petty & Brinol, 2008; Vraga & Bode, 2018; Vraga et al., 2019).

With regards to direct effects, the study confirmed most of the theorized relationships

among the correction strategy, attention, message credibility, and HPV misperceptions. First, the humorous correction produced more attention to the misinformation text (not the image) as well as the correction image (not the text) than the non-humorous correction (per RQ1 and RQ2), confirming the engaging role that humor can play and extending it to the context of correction to misinformation. In contrast, however, the non-humorous correction received higher credibility ratings than the humorous correction (per RQ3b), suggesting that credibility and attention to the corrections are not fully aligned, which explain the lack of direct effect of correction strategy on the credibility of the misinformation or HPV misperceptions (per RQ3a and RQ4).

In terms of indirect effects, the study findings revealed several interesting patterns (RQ 5). As hypothesized, credibility ratings of the misinformation were positively associated with HPV misperceptions (H3) while credibility ratings for the correction are negatively associated with HPV misperceptions (H4). Combined with a significant negative association between attention to the correction image (but not text) and misinformation credibility (supporting H2a but not H2b), the effects of correction strategy on HPV misperceptions were fully mediated. Therefore, increased attention to the *correction image* for the humorous correction—rather than the misinformation text or image itself—reduced credibility assessments of the misinformation tweet and thus reduced HPV misperceptions. This may not be surprising in this context; the text of the correction messages was held constant across conditions, minimizing the potential for differences to occur. However, it may also indicate the power of visuals on social media, which are shown to boost attention (e.g., Vraga et al., 2016), and recognition of the importance of visual cues as part of correction strategies overall (Bolsen et al., 2019; Nyhan & Reifler, 2019).

The humorous correction also increased attention to the text of the *misinformation* post, which was also consistent across conditions. The humorous parallel argumentation format,

which shows the logical flaw in the misinformation through comparison to an extreme application of that flaw (Cook et al., 2018), may have encouraged people to re-examine the misinformation message through the lens of the rhetorical flaws the correction exposed in a way that did not occur for the non-humorous logic-based correction. Future research should disentangle whether it is the use of humor or the structure of the logical rebuttal (parallel argument versus deconstruction of rhetorical flaws) that produces this effect.

Most importantly, our study found that attention was only the first step in the mediated pathways to reduce HPV misperceptions. In other words, people in the humorous correction paid more attention to the correction image (i.e., the cartoon) than people in the non-humorous correction—and this increased attention to the correction image reduced the credibility of the misinformation tweet, which then contributed to lower HPV misperceptions. In contrast, when the correction message was seen as *more* credible (as in the case of the non-humor correction), people reduced their HPV misperceptions without an indirect pathway via attention. These results not only align with research on the importance of credibility in persuasion and correction (Lewandowsky et al., 2012; Petty & Brinol, 2008; Pornpitakpan, 2004), but extend our understanding into how attention *shapes* these credibility assessments, confirming its central role preceding information processing (Lang, 2000; 2006; Petty & Cacioppo, 1986).

Based on our study findings, those interested in responding to misinformation on social media may face competing goals. The humorous parallel argumentation correction generated more attention—and future research should also investigate whether that leads people to share the correction more widely. But the non-humorous argument deconstruction correction was seen as more credible—and this credibility reduced HPV misperceptions. Therefore, responding to misinformation may require scholars and practitioners to consider what would be the right

amount of humor appeal to ensure both attention and credibility goals are met.

This study acknowledges its limitations. First, our two corrections were not perfectly parallel: the humorous correction used parallel argumentation while the non-humorous correction used argument deconstruction. We deemed this necessary to maximize the difference in humor between the two conditions, as these two communication approaches were particularly suited to humor or non-humor, enhancing realism. Additionally, the humorous image contained more words than the non-humorous image, which may have contributed to greater attention. Future studies could use additional conditions to untangle these effects, perhaps by incorporating humor (or not) into the text rather than the visual images as done here.

Second, while we made conscious efforts to mimic actual information environments as much as possible, the nature of an eye tracking experiment required the study to be conducted in a controlled lab setting, which might affect ecological validity. Additionally, our study included one correction tweet but multiple corrections can have a social norming effect on the credibility of an original message and improve the likelihood that the replies will reduce misperceptions (Lewandowsky et al., 2019; Vraga & Bode, 2017).

Finally, we focus on total fixation duration as our measure of attention, but future research should incorporate other measures like message recall or comprehension (e.g., thought-listing) to enhance our understanding of what attention signals. Although our attention measures occurred before posttest credibility and misperceptions, which helped with establishing the temporal order assumption for mediation analysis (MacKinnon et al., 2007), our data were cross-sectional in nature. Future research should test if the effects found in this study would persist over time—and if so, how long—by using a longitudinal design.

Despite the limitations, this study provides unique theoretical, methodological, and

practical implications with regards to health misinformation and correction on social media. This is the first study to empirically investigate the potential mediating roles of message attention and credibility on social media to reduce HPV misperceptions, reinforcing the ELM's explanatory power. Eye tracking is an unobtrusive method to objectively investigate visual attention, and the implementation of the eye tracking measures can complement and supplement subjective self-reported data. Eye tracking can also capture conscious and subconscious visual attention among individuals and reveal subtle and nuanced behavioral patterns (Holmqvist et al., 2011). With eye tracking technology, this study is able to speak to the importance of message attention as a mechanism by which correction functions.

Our study offers practical insights for scholars, educators, policy makers, and other public health professionals interested in developing effective health communication campaigns to address misinformation on social media. Both humor and non-humor logic-based correction strategies can produce observational correction, providing a new avenue beyond expert or factual correction that extends to issues where expert consensus is less robust. Attention and credibility are both important for successful correction, but it can be hard to leverage both in a single correction. Future research should explore whether it is possible for communicators to design refutation interventions that can attract both attention and credibility in reducing misinformation.

References

Becker, A. B., & Waisman, D. J. (2013). From funny features to entertaining effects: Connecting approaches to communication research on political comedy. *Review of Communication*, *13*(3), 161-183.

Bode, L., & Vraga, E. K. (2015). In related news, that was wrong: The correction of misinformation through related stories functionality in social media. *Journal of Communication*, *65*(4), 619-638.

Bolsen, T., Palm, R., & Kingsland, J. T. (2019). Counteracting climate science politicization with effective frames and imagery. *Science Communication*, *41*(2), 147-171.

Bore, I. L. K., & Reid, G. (2014). Laughing in the face of climate change?: Satire as a device for engaging audiences in public debate. *Science Communication*, *36*(4), 454-478.

Brewer, P. R., & McKnight, J. (2017). "A statistically representative climate change debate": Satirical television news, scientific consensus, and public perceptions of global warming. *Atlantic Journal of Communication*, *25*(3), 166-180.

Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., Quinn, S. C., & Dredze, M. (2018). Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *American Journal of Public Health*, *108*(10), 1378-1384.

Campo, S., Askelson, N. M., Spies, E. L., Boxer, C., Scharp, K. M., & Losch, M. E. (2013). "Wow, That Was Funny": The value of exposure and humor in fostering campaign message sharing. *Social Marketing Quarterly*, *19*(2), 84-96.

Compton, J. (2018). Inoculation against/with political humor. In J. C. Baumgartner & A. B. Becker (Eds.), *Political humor in a changing media landscape: A new generation of research* (pp. 95-114). Lexington Books.

Cook, J., Lewandowsky, S., & Ecker, U. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PloS One*, *12*(5): e0175799. https://doi.org/10.1371/journal.pone.0175799

Cook, J., Ellerton, P., & Kinkead, D. (2018). Deconstructing climate misinformation to identify reasoning errors. *Environmental Research Letters*, *13*(2), 024018.

Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet information credibility. *Journalism & Mass Communication Quarterly*, *77*(3), 515-540.

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.

Guidry, J. P., Carlyle, K., Messner, M., & Jin, Y. (2015). On pins and needles: How vaccines are portrayed on Pinterest. *Vaccine*, *33*(39), 5051-5056.

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.

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

Hameleers, M., & van der Meer, T. G. (2019). Misinformation and polarization in a high-choice media environment: How effective are political fact-checkers? *Communication Research*, 47(2), 227-250.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Houts, P. S., Doak, C. C., Doak, L. G., & Loscalzo, M. J. (2006). The role of pictures in improving health communication: A review of research on attention, comprehension,

recall, and adherence. *Patient Education and Counseling*, *61*(2), 173-190.

Jolley, D., & Douglas, K. M. (2014). The effects of anti-vaccine conspiracy theories on vaccination intentions. *PloS One*, *9*(2), e89177.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*(4), 441-480.

Juthe, A. (2009). Refutation by parallel argument. *Argumentation*, *23*(2), 133-169.

Kim, S. C., Namkoong, K., Fung, T., Heo, K., & Gunther, A. (2018). Understanding public opinion change of HPV vaccination controversy: Effects of exemplification and the mediating role of projection. *Health Education, 118*(5), 402-412.

King, A. J., Bol, N., Cummins, R. G., & John, K. K. (2019). Improving visual behavior research in communication science: An overview, review, and reporting recommendations for using eye-tracking methods. *Communication Methods and Measures, 13(3), 149-177.*

Lang, A. (2000). The limited capacity model of mediated message processing. *Journal of Communication*, *50*(1), 46-70.

Lang, A. (2006). Using the limited capacity model of motivated mediated message processing to design effective cancer communication messages. *Journal of Communication*, *56*, S57-S80.

Lazard, A., & Atkinson, L. (2015). Putting environmental infographics center stage: The role of visuals at the elaboration likelihood model's critical point of persuasion. *Science Communication*, *37*(1), 6–33.

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.

Lewandowsky, S., Cook, J., Fay, N., & Gignac, G. E. (2019). Science by social media: Attitudes towards climate change are mediated by perceived social consensus. *Memory & Cognition, 47*(8), 1445-1456.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58*, 593-614.

Margolin, D. B., Hannak, A., & Weber, I. (2018). Political fact-checking on Twitter: When do corrections have an effect? *Political Communication, 35*(2), 196-219.

McGraw, A. P., Schiro, J. L., & Fernbach, P. M. (2015). Not a problem: A downside of humorous appeals. *Journal of Marketing Behavior, 1,* 187-208.

Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R., & McCann, R. (2003). Bringing the concept of credibility into the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. *Communication Yearbook, 27*, 293-335.

Moyer-Gusé, E., Robinson, M. J., & Mcknight, J. (2018). The role of humor in messaging about the MMR vaccine. *Journal of Health Communication, 23*(6), 514-522.

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.

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

Nyhan, B., & Reifler, J. (2019). The roles of information deficits and identity threat in the prevalence of misperceptions. *Journal of Elections, Public Opinion and Parties*, *29*(2), 222-244.

Olsen, A. (2012). *The Tobii I-VT fixation filter.* Tobii Technology.

Parker, K. A., Ivanov, B., & Compton, J. (2012). Inoculation's efficacy with young adults' risky behaviors: Can inoculation confer cross-protection over related but untreated issues? *Health Communication, 27*(3), 223-233.

Petty, R. E., & Brinol, P. (2008). Persuasion: From single to multiple to metacognitive processes. *Perspectives on Psychological Science, 3*(2), 137-147.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, (Vol. 19, pp. 87-136). Academic Press.

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, *34*(2), 243-281.

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

Skurka, C., Niederdeppe, J., Romero-Canyas, R., & Acup, D. (2018). Pathways of influence in emotional appeals: Benefits and tradeoffs of using fear or humor to promote climate change-related intentions and risk perceptions. *Journal of Communication, 68*, 169-193.

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.

Stroud, N. J. (2017). Attention as a valuable resource. *Political Communication, 34*(3), 479-489.

Sülflow, M., Schäfer, S., & Winter, S. (2019). Selective attention in the news feed: An eye-tracking study on the perception and selection of political news posts on Facebook. *New Media & Society, 21*(1), 168-190.

Thorson, E. (2016). Belief echoes: The persistent effects of corrected misinformation. *Political*

Communication, 33(3), 460-480.

Tobii Pro (2019). *What is eye tracking?* https://www.tobiipro.com/blog/what-is-eye-tracking/

van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges, 1*(2), 1600008.

van der Meer, T. G., & Jin, Y. (2019). Seeking formula for misinformation treatment in public health crises: The effects of corrective information type and source. *Health Communication, 35*(5), 560-575.

Vraga, E. K., & Bode, L. (2017). Using expert sources to correct health misinformation in social media. *Science Communication, 39*(5), 621-645.

Vraga, E. K., & Bode, L. (2018). 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., Bode, L., & Troller-Renfree, S. (2016). Beyond self-reports: Using eye tracking to measure topic and style differences in attention to social media content. *Communication Methods and Measures, 10*(2-3), 149-164.

Vraga, E. K., Johnson, C. N., Carr, D. J., Bode, L., & Bard, M. T. (2014). Filmed in front of a live studio audience: Laughter and aggression in political entertainment programming. *Journal of Broadcasting & Electronic Media, 58*(1), 131-150.

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 and Politics.*

https://doi.org/10.1177/1940161220919082

Young, D. G. (2008). The privileged role of the late-night joke: Exploring humor's role in disrupting argument scrutiny. *Media Psychology, 11*(1), 119-142.

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.

Table 1. Zero-order corrections among variables of interest (n = 61).

Variables	1	2	3	4	5	6	7
1. Attention to Misinfo Text	1						
2. Attention to Misinfo Image	.01	1					
3. Attention to Correction Text	.64**	.27*	1				
4. Attention to Correction Image	.58**	.28*	.57**	1			
5. Misinfo Credibility	-.12	-.01	-.001	-.35**	1		
6. Correction Credibility	.03	.04	.02	.09	-.44**	1	
7. HPV Misperceptions	-.004	.12	.13	-.01	.61**	-.49**	1

Note. *p < 0.05, **p < .01.

Table 2. Direct and indirect effects of correction strategy on HPV misperceptions through message attention and misinformation credibility (n = 58).

	First Mediator				Second Mediator	Final Outcome
Direct Effects	Attention to Misinfo Text	Attention to Misinfo Image	Attention to Correction Text	Attention to Correction Image	Misinfo Credibility	HPV Misperceptions
Pretest Score of HPV Misperceptions	-.30 (.57) [-1.44, .84]	.42 (.41) [-.39, 1.24]	.51 (.67) [-.84, 1.86]	-.10 (.58) [-1.25, 1.05]	.56 (.14)*** [.29, .84]	.63 (.10)*** [.43, .84]
Non-humor Correction	-2.20 (.97)* [-4.15, -.25]	.56 (.58) [-.60, 1.73]	-2.41 (1.27) [-4.96, .13]	-2.25 (.99)* [-4.23, -.28]	-.25 (.36) [-.98, .48]	-.13 (.18) [-.49, .24]
Attention to Misinfo Text	-	-	-	-	.03 (.06) [-.10, .15]	-.004 (.04) [-.08, .07]
Attention to Misinfo Image	-	-	-	-	.04 (.10) [-.16, .24]	.004 (.06) [-.11, .12]
Attention to Correction Text	-	-	-	-	.06 (.04) [-.02, .13]	-.01 (.03) [-.06, .05]
Attention to Correction Image	-	-	-	-	-.21 (.06)** [-.33, -.09]	.04 (.04) [-.05, .13]
Misinfo Credibility	-	-	-	-	-	.30 (.09)** [.12, .48]
R2 %	9.69 %	3.91 %	9.04 %*	9.33 %	33.09 %***	70.61 %***
Indirect Effects	Non-humor Correction → Attention to Correction Image → Misinformation Credibility → HPV Misperceptions: b = .14 (.08)* [.02, .31]					

Note. Coefficients are unstandardized betas (b). Standard errors (SEs) are reported in parenthesis. Confidence intervals are reported in bracket. Only significant indirect effects are reported. *p < 0.05, **p < 0.01, ***p < 0.001

Table 3. Direct and indirect effects of correction strategy on HPV misperceptions through message attention and correction credibility (n = 58).

	First Mediator				Second Mediator	Final Outcome
Direct Effects	Attention to Misinfo Text	Attention to Misinfo Image	Attention to Correction Text	Attention to Correction Image	Correction Credibility	HPV Misperceptions
Pretest Score of HPV Misperceptions	-.30 (.57) [-1.44, .84]	.42 (.41) [-.39, 1.24]	.51 (.67) [-.84, 1.86]	-.10 (.58) [-1.25, 1.05]	-.29 (.18) [-.65, .06]	.73 (.11)*** [.51, .94]
Non-humor Correction	-2.20 (.97)* [-4.15, -.25]	.56 (.58) [-.60, 1.73]	-2.41 (1.27) [-4.96, .13]	-2.25 (.99)* [-4.23, -.28]	.87 (.38)* [.12, 1.63]	.03 (.20) [-.38, .44]
Attention to Misinfo Text	-	-	-	-	-.01 (.08) [-.16, .15]	.001 (.04) [-.09, .09]
Attention to Misinfo Image	-	-	-	-	-.03 (.08) [-.19, .13]	.01 (.06) [-.10, .12]
Attention to Correction Text	-	-	-	-	.02 (.05) [-.08, .12]	.02 (.03) [-.05, .09]
Attention to Correction Image	-	-	-	-	.06 (.06) [-.06, .18]	-.01 (.04) [-.10, .08]
Correction Credibility	-	-	-	-	-	-.27 (.08)*** [-.42, -.11]
R2 %	9.69 %	3.91 %	9.04 %*	9.33 %	17.86 %*	67.10 %***
Indirect Effects	Non-humor Correction → Correction Credibility → HPV Misperceptions: b = -.23 (.12)* [-.52, -.03]					

Note. Coefficients are unstandardized betas (b). Standard errors (SEs) are reported in parenthesis. Confidence intervals are reported in bracket. Only significant indirect effects are reported. *p < 0.05, **p < 0.01, ***p < 0.001

Figure 1. A conceptual model.

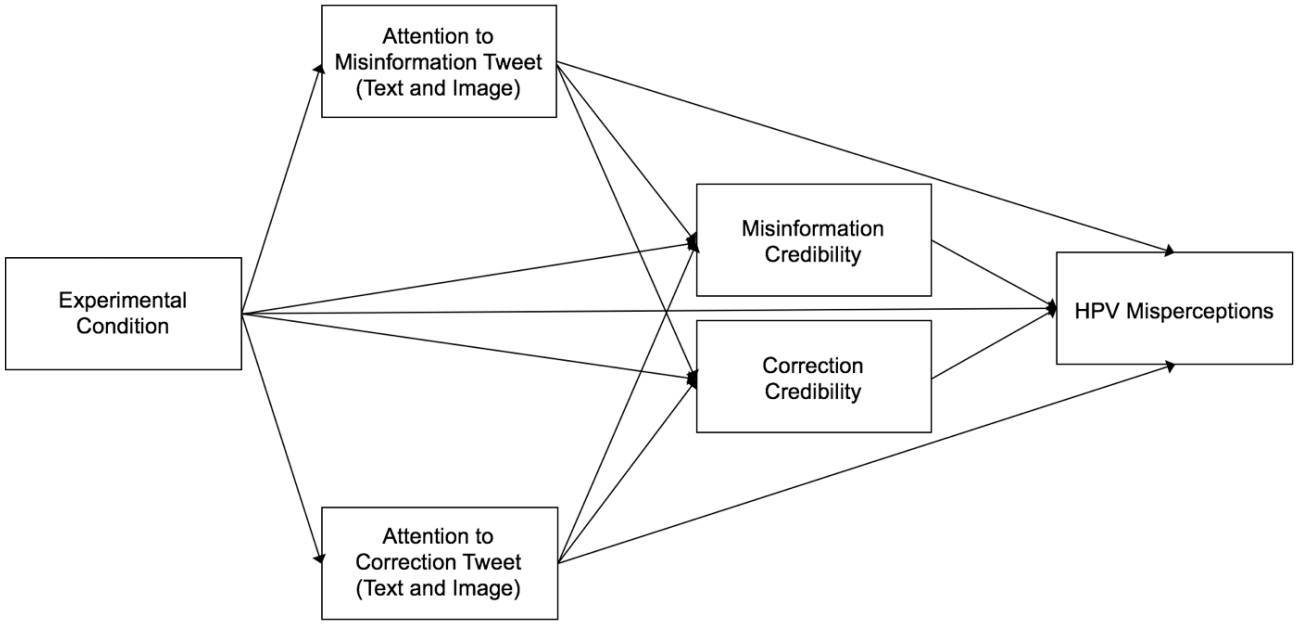

Figure 2. AOIs for humorous and non-humorous correction tweets