MIS Quarterly

RESEARCH NOTE

CAN FACT-CHECKING INFLUENCE USER BELIEFS ABOUT MISINFORMATION CLAIMS: AN EXAMINATION OF CONTINGENT EFFECTS[1]

Anol Bhattacherjee
School of Information Systems and Management, University of South Florida
Tampa, FL U.S.A. {abhatt@usf.edu}

Prior research has suggested that corrective fact-checking has inconsistent effects on beliefs about online misinformation claims. This study attempts to explain this inconsistency using three contingent factors—claim-source credibility, fact-checker credibility, and attitude strength—which respectively relate to three key parties in the fact-checking process: the source of a misleading claim, the fact-checker, and the user evaluating the fact-check. I hypothesize the interplay between these factors, which is tested using two online experiments on COVID-19-related misinformation with over 900 participants. Multilevel analysis of pretest-posttest, repeated measures data supports the hypothesized moderating effects and offers additional insights about how these effects vary between earlier versus later phases of misinformation cycles. The paper concludes with a discussion of contributions to research and practice.

Keywords: Misinformation, fact-checking, source credibility, attitude strength, experiment

Motivation for the Study

According to a 2019 Ipsos poll of over 25,000 online users in 25 countries, 86% of users have been exposed to disinformation or misinformation on social media platforms, news websites, YouTube, or television, and almost 9 out of 10 users initially believed such misinformation to be true (Simpson, 2019). Disinformation, also called fake news, refers to information that is known to be false but is spread intentionally to mislead others, while misinformation refers to false or out-of-context information that is mistakenly believed to be true and spread without the intent to deceive or mislead (Hernon, 1995). While disinformation has been used since ancient Greece and Rome to manipulate public opinion (Grant, 1995), it has gained considerable prominence in recent years in political campaigns, wars, and pandemic management.

To counter the growing scourge of disinformation and misinformation, news agencies, news aggregators, and social media platforms have increasingly turned to fact-checking to correct inaccurate claims and counterclaims from politicians, pundits, and others. Fact-checking is performed today by over 180 fact-checking organizations across 60 countries, which constitute the International Fact-Checking Network (Stencel, 2019). This network includes fact-checking units of major news organizations, such as AP Fact Check from the Associated Press, Fact Checker from the *Washington Post*, and Reality Check from the BBC, as well as independent fact-checkers such as Snopes, Politifact, and FactCheck.

But does fact-checking indeed help correct user perceptions of misinformation? Prior research has provided mixed answers to this question. Some studies have reported that fact-checking corrects falsely held beliefs (Weeks & Garrett, 2014; Pingree et al., 2014; Wood & Porter, 2019), while others have found no such effect (Jarman, 2016; Thorson, 2016; Moravec et al., 2019) and still others have suggested that fact-checking may even backfire by causing people to embrace their false beliefs more strongly rather than reject them (e.g., Nyhan & Riefler, 2010). Meta-analyses of fact-checking research have also reported varying results: Chan et al. (2017) found that fact-checking has a large effect on user perceptions, Walter and

[1] Shuk Ying Ho was the accepting senior editor for this paper. Chuan Hoo Tan served as the associate editor.

Murphy (2018) reported a small to moderate effect, Walter et al. (2019) found a small effect on political beliefs, and Walter and Tukachinsky (2020) observed a small *negative* effect.

In light of these inconsistent findings in the fact-checking literature, the research question of interest to this study is: *What contingent factors, if any, influence users' believability perceptions of online misinformation following corrective fact-checking, and how?* To answer this research question, I draw upon two complementary perspectives from the psychology literature: source credibility and attitude strength. Source credibility focuses on message sources (e.g., their trustworthiness) while attitude strength focuses on message recipients (i.e., the resilience of their attitude). I further distinguish between two types of sources, primary (claim sources) and secondary (fact-checkers), and examine the joint effects of claims sources, fact-checkers, and claims recipients on user beliefs following a corrective fact-check. I hypothesize that fact-checking efficacy is shaped by the interplay between claim-source credibility, fact-checker credibility, and users' attitude strength, and test these hypotheses using two within-subject online experiments involving COVID-19-related misinformation. Multilevel analysis of pretest-posttest, repeated measures data confirmed that claim-source credibility, fact-checker credibility, and attitude strength moderate the extent to which fact-checking influences users' believability perceptions of misinformation.

Given that the primary goal of fact-checking is to correct erroneous claims, this study examines corrective fact-checking only and excludes fact-checking that affirms correct information. Following the context-specific theorizing approach advanced by Hong et al. (2014), this study builds on general theories that are relevant to a problem domain, contextualizes these theories by identifying specific constructs relevant to the specific problem, and examines the interplay between these theories and the technology artifact (e.g., "false" flags by fact-checkers) in the form of interaction or moderating effects.

The theoretical contributions of this study are its explication of three key factors that shape the efficacy of corrective fact-checking and its elaboration of the pathways by which these factors shape user beliefs. In doing so, this study contributes to a preliminary theory of fact-checking. To the best of my knowledge, no study has yet examined source credibility effects in a fact-checking context and although strong user attitudes are known to resist fact-checking ("confirmation bias"), little is known about the interplay between user attitudes and claim-source or fact-checker credibility during the fact-checking process. This study also informs the source credibility literature on the joint effects of contradicting information sources. For practice, this study informs fact-checkers about how they are perceived by the user community and what they can do to help improve their credibility.

Related Literature

The central question in fact-checking research is to what extent fact-checking corrects falsely held beliefs. Following a meta-analysis of 52 samples from 20 studies drawn from eight research reports, Chan et al. (2017) reported that fact-checking has a large effect on user beliefs about misinformation (Hedges' $g = 0.88$; Pearson's correlation $r = 0.40$). However, Walter and Murphy's (2018) meta-analysis of 65 samples from 45 studies found a moderate effect on health beliefs (correlation $r = 0.27$) and a small effect on political beliefs ($r = 0.15$) and marketing beliefs ($r = 0.18$). This study concluded that political disinformation, given its partisan nature, is resistant to fact-checking, while less partisan health beliefs may be more amenable to fact-checking. Walter et al.'s (2019) meta-analysis of 30 studies from 20 research reports found a small effect (Cohen's $d = 0.29$; $r = 0.13$) on political beliefs. However, Walter and Tukachinsky's (2020) meta-analysis of 32 studies from 21 research reports observed a weak *negative* effect ($r = -0.05$), suggesting that misinformation may continue to shape one's beliefs even after being exposed to corrections.

While meta-analyses have provided inconsistent estimates of fact-checking's average effect size, individual effects reported in empirical studies are wildly inconsistent. Some studies have reported that fact-checking corrects user perceptions of misinformation claims (Weeks & Garrett, 2014; Wood & Porter, 2019), increases participants' belief accuracy and political cynicism (Pingree et al., 2014), and makes people less convinced of fake news stories (Porter et al., 2018). Other studies have observed that people tend to maintain stable attitudes toward misinformation, even after it is debunked (Thorson, 2016); the presence of a fake news flag does not change students' perceptions of truth although it increases their cognitive activity and motivates them to spend more time in reviewing misinformation (Moravec et al., 2019); and partisan users are especially resistant to fact-checking (Nyhan & Riefler, 2010). Nyhan and Reifler (2010) even noted that fact-checking may cause some people to embrace false beliefs more strongly rather than abandon them, which they called the "backfire effect."

Researchers have also examined attributes of the fact-checking process and individual differences among users as predictors of fact-checking efficacy. Studies on the fact-checking process attributes have shown that truth scales (e.g., mostly false / partially false / mixed / partially true / mostly true) tend to convince people more than binary scales (true/false) for nonpolitical claims but not for political claims (Amazeen et al., 2018); stronger ratings ("mostly false") are more effective than weaker ratings ("partially false") (Jarman, 2016); corrective fact-checks are more effective than confirming fact-checks (Fridkin et al., 2015); and the volume of corroborating evidence (Ecker et al., 2015), type of

evidence, such as video vs. text (Young et al. 2017), and contextual information and pictures (Garrett et al., 2013) influence fact-checking outcomes. Studies on individual differences have indicated that users with stronger political affiliation with the claim source are less influenced by fact-checking (Jarman, 2016); those with strong preexisting beliefs resist contradictory fact-checks even in the face of contrary evidence (Walter et al., 2019; Moravec et al., 2019); and gender, tolerance of negativity, and political sophistication also influence fact-checking outcomes (Fridkin et al., 2015).

Much of the above research cannot quite explain why fact-checking works in some instances but not in others. This is presumably because of its focus on the main effects of predictors rather than the moderating effects. This paper explores the moderating effects of three key factors: claim-source credibility, fact-checker credibility, and users' attitude strength. The next section draws upon source credibility and attitude strength as theoretical lenses to examine how these three factors jointly influence corrective fact-checking efficacy.

Theoretical Background and Hypotheses

Fact-checking can be viewed as a persuasion process in which fact-checkers attempt to persuade users to change their previously held beliefs about misinformation claims. Hence, theories of persuasion are relevant to understanding fact-checking. I examine two such theoretical perspectives to identify constructs salient to corrective fact-checking.

Source Credibility

Theories of persuasion have long held that source credibility is a key determinant of message effectiveness (Hovland & Weiss, 1951; Hovland et al., 1953). Source credibility is defined as the extent to which one perceives an information source to be believable, competent, and trustworthy (Petty & Cacioppo, 1986). Source credibility theory states that people are more likely to be persuaded by a message if they see the message as coming from a credible or trustworthy source (Hovland & Weiss, 1951; Hovland et al., 1953). Through a process of trust transference, one's trust in a credible message source is transferred to claims made by that source. Conversely, claims from a less credible source are less likely to be believed.

While prior research has confirmed the importance of source credibility in the formation of user beliefs and attitudes (e.g., Hovland et al., 1953; McGinnies, 1973), it is still unclear whether such beliefs will persist following a corrective fact-check. The corrective fact-checking context is characterized by two contradicting sources of information: the original claim source, referred to as the *primary source*, and the fact-checker disputing that claim, referred to as the *secondary source*. While a highly credible primary source may increase the believability of misinformation claims coming from that source, a high-credibility secondary source (fact-checker) disputing the veracity of those claims may similarly decrease users' believability perceptions of fact-checked claims.

If users view a fact-checker as being highly credible, they are more likely to believe its fact-checks; hence, their believability perceptions of a misinformation claim will be lower following a corrective fact-check, relative to their perceptions before the fact-check. Note that affirmative fact-checks may increase the believability of such claims, but such fact-checking is beyond the scope of this study. While the credibility of news sources has been examined in the source credibility literature, fact-checker credibility (i.e., the credibility of a secondary source in the presence of a primary source) is novel to both source credibility and fact-checking research. To test for this effect, based on the preceding rationale, I propose:

H1: *Fact-checker credibility has a negative effect on users' believability perceptions of online claims following corrective fact-checks.*

When claim-source credibility and fact-checker credibility are concurrently in effect, how do they jointly impact users' believability perceptions? If a claim from a less credible claim source is contradicted by a fact-check from a highly credible fact-checker, the higher credibility of the secondary source may motivate users to update (lower) their pre-fact-check beliefs about that claim. However, if the same claim is contradicted by a less credible fact-checker, users may discount the fact-check and not change their pre-fact-check beliefs as expected. Hence, one might expect an interaction between claim-source credibility and fact-checker credibility such that claim-source credibility can partially mitigate the negative effects of fact-checking credibility on users' beliefs following a corrective fact-check. Hence, I hypothesize:

H2: *The negative effect of fact-checker credibility on users' believability perceptions of online claims following a corrective fact-check is higher for claim sources with low credibility than for those with high credibility.*

Attitude Strength

Research on human attitudes suggests that attitudes vary in strength across the population and this strength determines the extent to which people are susceptible to influence by external information (Howe & Krosnick, 2017). Attitude strength is defined as the power of attitude to resist change (Krosnick &

Petty, 1995), and is conceptualized as how strongly users feel about a claim, how certain they are about their attitudes about the claim, and how important they perceive the claim to be (Krosnick & Abelson, 1992). Strong attitudes tend to persist over time, even after being refuted by evidence, and are more predictive of longer-term human behavior than weak attitudes. Seemingly immovable public attitudes have often been blamed for the persistence of many social evils such as slavery and racism, and significant societal changes occurred only after these strong attitudes were weakened.

The underlying psychological processes by which strong attitudes resist change are *selective exposure* and *selective elaboration* (Howe & Krosnick, 2017). When existing attitudes about an online claim are at odds with new information such as a corrective fact-check, this contradiction creates cognitive dissonance (Festinger, 1957), which people seek to reduce by deliberately seeking information that supports preexisting attitudes, limiting exposure to contradictory information (selective exposure), and overweighting arguments that favor preexisting attitudes and underweighting evidence to the contrary (selective elaboration) (Brannon et al., 2007). The resulting evaluation, also called confirmation bias or motivated reasoning (Nickerson, 1998), is biased in favor of the initially held beliefs, even if those beliefs are incorrect in the first place. Hence, if people hold strong initial attitudes about a misinformation claim, they will look for reasons to reject corrective fact-checks; thus, such fact-checking may only work for claims about which weak initial attitudes are held.

While the effect of preexisting attitudinal biases is well-documented in the fact-checking literature (e.g., Moravec et al., 2019) and is therefore not hypothesized in this study, less is known about how attitude strength influences beliefs following exposure to claim sources and fact-checkers of varying credibility. Strong attitudes tend to inhibit belief changes following corrective fact-checks, as does high claim-source credibility. In tandem, these two effects may reinforce each other, such that the mean positive effect of claim-source credibility on users' believability perceptions is amplified for users with strong attitudes and attenuated for users with weaker attitudes. In contrast, fact-checker credibility and attitude strength have opposite effects on one's beliefs. Hence, attitude strength may negatively moderate the effect of fact-checker credibility on users' believability perceptions after corrective fact-checks, such that this effect is weaker for users with strong attitudes than for those with weak attitudes. These expectations lead to the following hypotheses:

H3: *The positive effect of claim-source credibility on users' believability perceptions of online claims following a corrective fact-check is higher for users with strong attitudes than for those with weak attitudes.*

H4: *The negative effect of fact-checker credibility on users' believability perceptions of online claims following a corrective fact-check is higher for users with weak attitudes than for those with strong attitudes.*

The four hypotheses are illustrated in Figure 1. H2 and H4 depict negative interaction or moderating effects, while H3 suggests a positive moderating effect. The next section describes two online experiments designed to test these hypotheses in the context of misinformation related to COVID-19 (coronavirus).

Study One

Methods

This experiment was conducted in 2020, during an early phase of the COVID-19 pandemic. The COVID-19 context was chosen for this study because of its contemporary relevance, the ubiquity of COVID-related misinformation, and the widespread corrective fact-checking of such misinformation by many fact-checkers. Participants in this experiment were recruited using Amazon Mechanical Turk (MTurk), with the sample restricted to U.S. adults with a human intelligence task (HIT) approval rating of over 85% to ensure high-quality responses. MTurk samples have previously been used in fact-checking research (e.g., Pennycook & Rand, 2019; Pennycook et al., 2020; Moravec et al., 2020).

The experiment employed a pretest-posttest, counterbalanced, repeated-measures design with within-subject treatments (see Figure 2). This design was chosen to control for participant-level and claim-level variations and is fairly typical of fact-checking studies (e.g., Pennycook & Rand, 2019). In the pretest phase, participants were introduced to two claim sources (Reuters and BuzzFeed, as manipulations of high and low credibility claim sources respectively) and two fact-checkers (FactCheck and HoaxSlayer, as high and low credibility fact-checkers), with their year of founding, number of employees, media awards, mode of operation, and website links. Participants were then asked four manipulation check questions to verify that they had carefully read the background information. Those who did not answer all four manipulation check questions correctly were dropped from the sample. Participants were also asked a series of questions to assess their credibility perceptions of each claim source and each fact-checker. They were then shown 10 COVID-19 claims in random order, and their prior exposure to and attitude strength toward each claim were measured using 5-point Likert scales, along with the perceived importance, relevance, interestingness, and believability of each claim. Measurement items and claims used are listed in the Appendix.

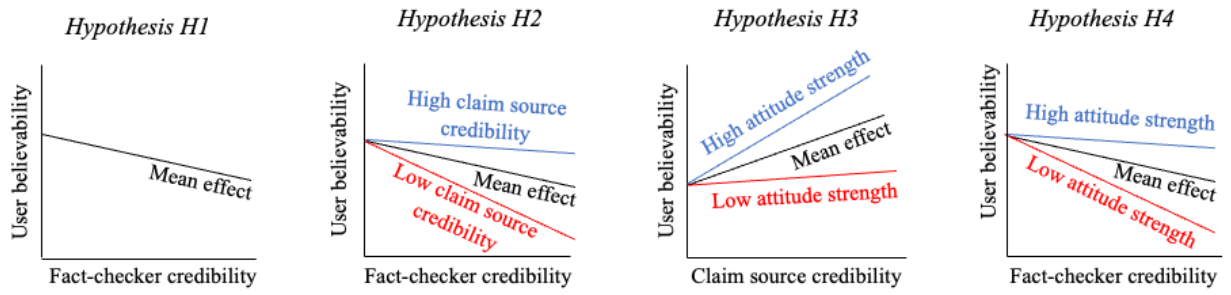

Figure 1. Hypotheses

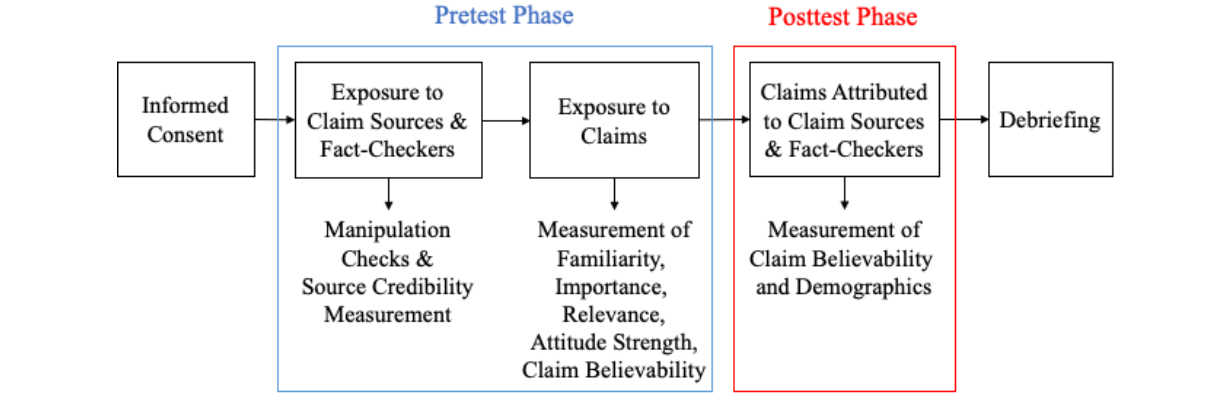

Figure 2. Experiment Design

In the posttest phase, the 10 COVID-19 claims were attributed randomly to one of the two claim source treatments (Reuters or BuzzFeed). Eight of the 10 claims were also randomly assigned to one of the two fact-checkers (FactCheck or HoaxSlayer) and assigned a "verified" (true) or "disputed" (false) rating, while the remaining two claims were marked "unchecked" and used as an experimental control. The unchecked claims provided some level of ecological validity in the study design since most online claims are not fact-checked. Participants were then asked to rate their post-treatment believability perceptions regarding each claim in light of the claim source, fact-checker, and fact-check rating of each claim. Given the focus on corrective fact-checking, all "verified" (true) claims were excluded from analysis. I also measured participants' willingness to read the complete article making the claim and their willingness to share the article with their social network, but this data was excluded from analysis given the triple-barreled nature of claim believability, reading intention, and sharing intention.

The choice of Reuters and BuzzFeed News as claim sources was motivated by a 2017 survey of 28 news sources by the University of Missouri's Reynold Journalism Institute (Kearney, 2017). This survey reported Reuters as one of America's most trusted, nonpartisan online news sources and BuzzFeed News as one of the least trusted. FactCheck was selected for the high credibility fact-checker treatment, given its reputation as a nonprofit, nonpartisan project of the Annenberg Public Policy Center of the University of Pennsylvania and its numerous awards for journalistic integrity from *TIME* magazine, the Society of Professional Journalists, and the International Academy of Digital Arts and Sciences. In contrast, HoaxSlayer was a one-person fact-checking operation, run from a home office in an outback town in Queensland, Australia, and was relatively unknown to the U.S. public.

The presentation of claims in the experiment was designed to mimic Google News, an online aggregator for customized news delivery. Google News dedicates a section of its web page to fact-checking, where it displays a list of claims made by different news sources, followed by fact-checks of those claims by independent fact-checkers like FactCheck or Snopes. Claim sources (news sites), fact-checker names, and fact-check ratings (true vs. false) are displayed on the web page, with hypertext links to fact-checkers' websites that

explain the basis for their rating. I did not include explanations for fact-check ratings in the experiment because there is little evidence that users actually read these explanations, and the effects of contextual information on fact-checking efficacy have been shown to be inconsistent (Garrett et al., 2013).

The claims used in this study were sourced from the U.S. Centers for Disease Control's (CDC) list of frequently asked COVID-19 questions (https://www.cdc.gov/coronavirus/2019-ncov/faq.html), where the CDC addressed common misperceptions about the pandemic. Because the study randomly labeled some true claims as false and some false claims as true, participants were debriefed on the ground truth of each claim, as evaluated by the CDC, at the end of the study. The study was approved by the Institutional Review Board at the author's university.

Measurement

The dependent variable in this study was participants' post-fact-check believability for each claim. Pre-fact-check and post-fact-check claim believability were each measured using a three-item scale adapted from Kim and Dennis (2020) and Moravec et al. (2020) that asked participants to rate how truthful, credible, and believable they considered each claim to be on 5-point semantic differential measures.

The independent variables were claim-source credibility, fact-checker credibility, and user attitude strength regarding each presented claim. Although claim-source credibility and fact-checker credibility were manipulated as treatments, I measured participants' perceptions of the same because behavioral treatments are only useful to the extent that they are perceived as such by participants. Drawing on prior research that describes trustworthiness and expertise (competence) as the two dimensions of source credibility (Petty & Cacioppo, 1986; Pornpitakpan, 2004), I measured credibility perceptions for each claim-source and fact-checker using three semantic differential items that questioned the extent to which participants considered each organization to be trustworthy and have the necessary expertise to do its job and asked them about their overall perceptions of the organization's credibility. Attitude strength was measured as how important the participant viewed each COVID-19 claim to be, by drawing on Moravec et al.'s (2019) measure.

Several additional variables were measured as control variables: participants' prior exposure to each claim (before this study), their perceived relevance and interestingness of each claim, and their age, gender, education, and other demographics. All measurement scales are listed in the Appendix.

Descriptive Analysis

Participants who did not spend at least five minutes on the task or did not correctly answer all four treatment manipulation check questions were dropped from the sample. The "five-minute rule" was based on a pretest that showed that it would take at least 7.5 minutes for a fast reader to read and answer all questions. This screening process led to a final sample of 434 participants and 4,290 observations that included 1,746 corrective ("disputed") fact-checks and 858 "unchecked" claims. The median time to complete the assigned task was 12.4 minutes, the median age of participants was 44 years, and the median educational level was "some level of college." A power analysis for a multiple regression model with a significance level of 0.05 and statistical power of 0.80 showed that the sample size of 434 would be able to detect effects as small as 0.03.

Cronbach's alphas for claim-source credibility and fact-checker credibility were 0.93 and 0.84, respectively. As shown in Table 1, the observed mean claim-source credibility for Reuters (high treatment) was 2.76 on a 5-point scale, compared to 1.94 for BuzzFeed News (low treatment). A one-sample *t*-test indicated that the difference in means was statistically significant (t = 20.61; $p < 0.001$). Similarly, the mean fact-checker credibility for FactCheck (high treatment) was 2.00, which was also significantly higher ($t = 7.93$; $p < 0.001$) than HoaxSlayer's 1.74 (low treatment). Hence, both treatment manipulations worked as intended. However, all four means were less than 3 (the neutral point on this scale), suggesting that on average, participants viewed both claim sources and fact-checkers with some degree of distrust. This observation is consistent with Gallup polls showing that only 36% of Americans trusted mass media in 2021, the second lowest level in history (Brenan, 2021). Particularly noteworthy was participants' lack of trust in FactCheck, despite its numerous industry awards and accolades.

Cronbach's alphas for pretest and posttest claim believability were 0.79 and 0.81, respectively. Mean pretest believability for all claims was 3.07, slightly exceeding the neutral value of 3, suggesting that in the absence of contrary evidence, people tend to believe claims of unknown veracity, which Levine and McCornack (1991) called the "truth bias." This means that believability dropped to 2.89 in the posttest phase, suggesting that participants revised their perceptions of COVID-19 claims following claim source and fact-checker exposure. This difference was statistically significant ($t = 12.65$; $p < 0.001$). For claims attributed to Reuters, mean claim believability fact-checked by FactCheck dropped from 3.04 (pretest) to 2.40 (posttest), while those fact-checked by HoaxSlayer dropped from 3.05 to 2.84. For BuzzFeed claims, mean claim believability fact-checked by FactCheck dropped from 3.06 (pretest) to 2.47 (posttest), while those fact-checked by HoaxSlayer dropped from 3.13 to 2.71. All of these changes were statistically significant at $p < 0.001$.

Table 1. Comparison of Means in Study 1			
Variable	**Mean (Std. dev.)**	**Mean (Std. dev.)**	***t*-statistic**
Claim-source credibility	RT: 2.756 (1.145)	BF: 1.935 (0.815)	20.617***
Fact-checker credibility	FC: 1.997 (0.787)	HS: 1.741 (0.590)	7.931***
Claim believability: All	Pretest (t_1): 3.074 (1.005)	Posttest (t_2): 2.796 (0.760)	18.240***
Claim believability: RT & FC	Pretest (t_1): 3.036 (1.053)	Posttest (t_2): 2.402 (0.761)	19.345***
Claim believability: RT & HS	Pretest (t_1): 3.046 (0.959)	Posttest (t_2): 2.838 (0.739)	6.200***
Claim believability: BF & FC	Pretest (t_1): 3.063 (1.030)	Posttest (t_2): 2.474 (0.687)	18.456***
Claim believability: BF & HS	Pretest (t_1): 3.126 (1.000)	Posttest (t_2): 2.709 (0.610)	12.274***

Note: RT: Reuters; BF: BuzzFeed; FC: FactCheck; HS: HoaxSlayer. *p < 0.05, **p < 0.01, ***p < 0.001

Regression Results

I ran three random-intercept, multilevel regression models to test the proposed hypotheses. Model 1 was a baseline model without the hypothesized variables of interest, Model 2 included the baseline model and hypothesized main effects, and Model 3 (full model) included the baseline model and hypothesized main and moderation effects (see Table 2). The dependent variable in each model was participants' post-fact-check claim believability. To ensure unbiased model estimates, pre-fact-check claim believability was included in each model. All models controlled for the fixed effects of participants (i = 1…n), claims (j = 1…10), and the order in which these claims were presented to participants (k = 1…10). I also included the ground truth of each claim as a predictor, but the effect of this variable could not be estimated and is thus not reported here. Model specification for the full model is shown in Equation (1).

$$\begin{aligned} &Claim\ Believability_{ijkt+1} = \\ &\beta_0 + \beta_1\ GroundTruth_{jt} + \beta_2\ PriorExposure_{ijkt} + \beta_3\ Relevance_{ijkt} + \beta_4\ Interestingness_{ijkt} + \\ &\beta_5\ ClaimBelievability_{ijkt} + \beta_6\ Attitude\ Strength_{ijkt} + \beta_7\ ClaimSourceCredibility_{ijkt} + \\ &\beta_8 FactcheckerCredibility_{ijkt} + \beta_9\ ClaimSourceCredibility_{ijkt} * FactcheckerCredibility_{ijkt} + \\ &\beta_{10}\ ClaimSourceCredibility_{ijkt} * AttitudeStrength_{ijkt} + \beta_{11}\ FactcheckerCredibility_{ijkt} * \\ &AttitudeStrength_{ijkt} + \sum_1^n \gamma_i\ Respondent_i + \sum_1^{10} \delta_j\ Claim_j + \sum_1^{10} \eta_k\ ClaimOrder_k + \varepsilon_{ijk} \end{aligned} \quad (1)$$

The baseline model had an adjusted R-squared value of 0.55, which increased to 0.63 in the main effects model and 0.66 in the moderation effects model. In the main effects model (Model 2), claim-source credibility had a positive effect on posttest believability for corrective fact-checks (β = 0.14, p < 0.001), as did attitude strength (β = 0.04, p < 0.001), consistent with confirmation bias expectations. Fact-checker credibility had a negative effect (β = -0.26, p < 0.001), supporting H1.

Based on the moderation effects model (Model 3), the interaction effect of claim-source credibility and fact-checker credibility on posttest claim believability were found to have a negative slope (β = -0.10, p < 0.001), while the moderation effect of attitude strength on the relationship between claim-source credibility and posttest believability was positive (β = 0.02, p < 0.001), supporting H2 and H3, respectively. However, the moderation effect of attitude strength on the relationship between fact-checker credibility and posttest believability was almost zero (β = -0.001, p > 0.01), failing to support H4.

I tested the robustness of the analysis using two additional models: (1) a full model where the perceived source and fact-checker credibility were replaced with the actual claim source (Reuters and BuzzFeed News) and fact-checker (FactCheck and HoaxSlayer) treatments as factor variables (Model 4), and (2) a Tobit model to account for the range-restricted nature of the dependent variable (Model 5). While Model 4 depicted a stronger main effect of fact-checker credibility, the effects in both models were largely consistent with those of Model 3, providing further support for the analysis.

One plausible reason for the weak main and moderation effects of attitude strength might have been the timing of this study, which was conducted during an early phase of the COVID-19 pandemic, when participants might not yet have had the opportunity to form strong attitudes about COVID-related claims due to the high level of uncertainty about the pandemic at that time. To examine whether these effects varied during the later stages of the pandemic, I conducted a second experiment, as described below.

Study Two

Two potential concerns about Study 1 were (1) its external validity (generalizability of inferences to other contexts) and (2) its ecological validity (generalizability to real-world settings). I addressed these concerns in a second study, which differed from Study 1 in several ways.

Table 2. Hypotheses Tests and Robustness Checks in Study 1

Variables	Model 1: Baseline model	Model 2: Model 1 + Main effects	Model 3: Model 2 + Moderation effects	Model 4: Objective indicator model	Model 5: Tobit model
Prior exposure	-0.036 (0.027)	-0.021 (0.025)	-0.027 (0.024)	-0.012 (0.025)	-0.019 (0.021)
Interestingness	-0.006 (0.018)	0.005 (0.016)	0.006 (0.015)	-0.008 (0.016)	0.007 (0.014)
Relevance	-0.057*** (0.018)	-0.036** (0.016)	-0.035** (0.016)	-0.035** (0.016)	-0.034* (0.014)
Pretest believability	0.511*** (0.014)	0.423*** (0.017)	0.425*** (0.017)	0.407** (0.018)	0.422*** (0.015)
Attitude strength	-	0.035*** (0.005)	-0.019 (0.012)	0.034*** (0.007)	-0.016 (0.011)
Claim-source credibility	-	0.140*** (0.015)	0.206*** (0.035)	-	0.247*** (0.030)
Fact-checker credibility	-	-0.263*** (0.022)	-0.007 (0.050)	-	0.067 (0.042)
Claim-source credibility × FC credibility	-	-	-0.104*** (0.016)	-	-0.124*** (0.014)
Attitude strength × Claim-source credibility	-	-	0.024* (0.003)	-	0.027*** (0.003)
Attitude strength × FC credibility	-	-	-0.001 (0.004)	-	-0.003 (0.004)
Reuters (vs. BuzzFeed)	-	-	-	0.099 (0.049)	-
FactCheck (vs. HoaxSlayer)	-	-	-	-0.197*** (0.049)	-
Reuters × FactCheck	-	-	-	-0.251*** (0.090)	-
Attitude strength × Reuters	-	-	-	0.012* (0.007)	-
Attitude strength × FactCheck	-	-	-	-0.0004 (0.007)	-
Intercept	0.599** (0.295)	0.851*** (0.274)	0.743*** (0.279)	0.814*** (0.271)	0.992*** (0.098)
Log(scale)	-	-	-	-	-0.865*** (0.018)
Observations	1,746	1,746	1,746	1,746	-
R-squared	0.664	0.728	0.747	0.724	-
Adjusted *R*-squared	0.547	0.632	0.658	0.627	-
F-statistic	5.700*** (*df* = 449;1296)	7.643*** (*df* = 452;1293)	8.390*** (*df* = 455;1290)	7.449*** (*df* = 455;1290)	-
Log-likelihood	-	-	-	-	-998.2 (*df* = 12)
Wald-statistic	-	-	-	-	3,198*** (*df* = 10)

Note: Standard errors are in parentheses. Fixed effects of participants, claims, and claim order not shown to conserve space. +$p < 0.10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

First, Study 2 was conducted in 2022, during a later phase of the pandemic. Study 1 was conducted when very little was known about the disease and participants may not yet have formed strong attitudes about COVID-19 policies. In contrast, when Study 2 was conducted, more information about the pandemic was available, and public attitudes about COVID-19 policies, such as masking and vaccines, had matured and hardened. Therefore, Study 2 allowed me to examine the external validity of the inferences across different levels of attitude strengths.

Second, Study 2 examined six COVID-19 claims that were different from those examined in Study 1 and were more relevant during the later phase of the pandemic (e.g., vaccine-related claims), which allowed me to examine the study's external validity across a different set of online claims.

Third, by the time Study 2 was conducted, HoaxSlayer had ceased operations, necessitating a new fact-checker. I replaced HoaxSlayer with a new and relatively unknown fact-checker called Health Feedback, which focuses solely on health-related fact-checking. Health Feedback employs as fact-checkers an unpaid community of scientists holding Ph.D. degrees who have published articles in top-tier peer-reviewed science journals and who fact-check claims in their domain of expertise. The names and profiles of these scientists are displayed on Health Feedback's website to convey their credibility to the public. In contrast, FactCheck employs paid fact-checkers with journalism degrees to fact-check claims across all domains. Hence, Health Feedback allowed me to examine the external validity of the study using a different fact-checker with a different fact-checking process and staff profile.

Fourth, to improve the ecological validity of the study, I employed a graded fact-checking scale, "mostly false," "partially false," and unchecked (control group), similar to that used by most fact-checkers, in contrast to a discrete scale (disputed, verified, or unchecked) in Study 1. Health Feedback rates its claims on a 5-point scientific credibility scale from -2 (very low) to 0 (neutral) to +2 (very high). Prior studies indicate

that fact-checking with graded scales tends to outperform binary (true/false) scales in correcting user perceptions (Amazeen et al., 2018). Further, given the study's focus on corrective fact-checking, I dropped the "true" (or verified) category from Study 1; I also dropped reading and sharing intention, which were unused in Study 1. Lastly, while non-fact-checked claims were explicitly labeled as "unchecked" in Study 1, these claims were left unlabeled in Study 2, similar to claims in the real world.

Fifth, another threat to ecological validity in Study 1 was participants' concurrent exposure to both claim sources and fact-checks. In reality, however, many users may see online claims from claim sources *before* they see their fact-checks. Prior exposure to claim sources may prime their attitudes against corrective fact-checks via an anchoring effect (Tversky & Kahneman, 1974). To isolate the effects of claim-source credibility and fact-checker credibility, Study 2 employed a three-phase design. Participants were exposed to the claims in Phase 1, to claim sources in Phase 2, and to fact-checkers in Phase 3. Participants' claim believability perceptions were measured at each of these three phases.

Lastly, in view of the weak effects of attitude strength in Study 1, this scale was expanded in Study 2 to include two additional items on how certain participants were about their attitudes about and how strongly they felt about the claims, based on Krosnick and Abelson's (1992) conceptualization of this construct.

Descriptive Analysis

Responses that were incomplete, did not pass treatment manipulation checks, and those from participants spending less than five minutes on the task were dropped from the sample. The final sample size was 527 participants and 3,162 observations. As seen in Table 3, mean claim-source credibility was 2.26 (out of 5.0) for the entire sample with a statistically significant difference between Reuters (2.64) and BuzzFeed (1.88). Mean fact-checker credibility was 2.21 for the entire sample, with a statistically significant difference between FactCheck (1.79) and Health Feedback (2.62). The mean credibility of Reuters, BuzzFeed, and FactCheck dropped from earlier (Study 1) to later (Study 2) phases of the pandemic. However, interestingly, the relatively unknown fact-checker Health Feedback had higher credibility than the more well-known FactCheck. This may be because Health Feedback's fact-checking staff, who are credentialed scientists in real life, were seen as being more credible in their domain of expertise than FactCheck's generic fact-checkers. The mean claim believability rating was 3.08 (out of 5) in Phase 1 (after claims exposure, but before source or fact-check exposure), 3.07 in Phase 2 (after source exposure but before fact-check exposure), and 2.93 in Phase 3 (after fact-check exposure). The difference-of-means test showed no significant change in mean claim believability from Phase 1 to Phase 2 ($t = 1.70$, $p > 0.05$), but a significant change from Phase 2 to Phase 3 ($t = 11.65$, $p < 0.001$), suggesting that the corrective fact-check had the expected effect. This change from the pretest to the posttest phase is similar to that in Study 1, suggesting that the two-phase design in Study 1 did not likely introduce artifactual variance, compared to the three-phase design in Study 2.

For claims attributed to Reuters, mean claim believability fact-checked by FactCheck changed from 3.18 in Phase 2 to 2.91 in Phase 3, while claims fact-checked by Health Feedback changed from 3.06 to 2.67, respectively. Mean believability for BuzzFeed claims fact-checked by FactCheck changed from 3.08 (Phase 2) to 2.86 (Phase 3), while those fact-checked by Health Feedback changed from 3.00 to 2.68. In each case, claim believability changes from Phase 2 to Phase 3 were statistically significant at $p < 0.05$.

Regression Results

Using Phase 3 believability as the dependent variable, I ran multilevel regression models, controlling for the fixed effects of participants, claims, and ordering of claims, similar to that in Study 1, but with one additional predictor: fact-checking strength (strong, weak, or none). Table 4 shows the results of this analysis. Adjusted *R*-squared values increased from 0.68 in Model 1 to 0.73 in Model 2 and 0.86 in Model 3. In the main effects model (Model 2), fact-checker credibility had a negative effect on posttest claim believability ($\beta = -0.12$, $p < 0.001$), supporting H1. Attitude strength had a positive effect, although, unlike Study 1, the effect of claim-source credibility was nonsignificant.

In Model 3, the interaction effect of claim-source credibility and fact-checker credibility on posttest claim believability was negative ($\beta = -0.02$, $p < 0.01$), supporting H2. Attitude strength moderated the association between claim-source credibility and posttest believability in a positive manner ($\beta = 0.06$, $p < 0.001$) and between fact-checker credibility and posttest believability in a negative manner ($\beta = -0.04$, $p < 0.001$), supporting H3 and H4, respectively.

Since participants in Study 2 viewed Health Feedback as more credible than FactCheck, I recoded Health Feedback and FactCheck as high and low credibility fact-checkers, respectively, in the objective indicator model (Model 4) to align this model with participants' self-reported perceptions in Model 3. Model 4's results are consistent with those of Model 3, as were those from the Tobit model for the range-restricted dependent variable, attesting to the robustness of the analysis.

Table 3. Comparison of Means in Study 2

Variable	Mean (Std. dev.)	Mean (Std. dev.)	Mean (Std. dev.)	*t*-statistic†
Claim-source credibility	All: 2.261 (0.761)	RT: 2.637 (0.749)	BF: 1.884 (0.561)	32.117***
Fact-checker credibility	All: 2.205 (1.005)	FC: 1.786 (0.657)	HF: 2.623 (1.114)	21.133***
Claim believability: All	t_1: 3.077 (1.326)	t_2: 3.069 (1.317)	t_3: 2.923 (0.986)	11.646***
Claim believability: RT & FC	t_1: 3.149 (1.294)	t_2: 3.183 (1.282)	t_3: 2.909 (0.984)	11.747***
Claim believability: RT & HF	t_1: 3.022 (1.294)	t_2: 3.056 (1.289)	t_3: 2.672 (0.820)	12.280***
Claim believability: BF & FC	t_1: 3.134 (1.343)	t_2: 3.080 (1.342)	t_3: 2.869 (0.738)	6.086***
Claim believability: BF & HF	t_1: 3.039 (1.365)	t_2: 2.996 (1.340)	t_3: 2.677 (0.597)	7.272***

Note: RT: Reuters; BF: BuzzFeed; FC: FactCheck; HF: Health Feedback. †Comparison of means between Columns 2 and 3. *p < 0.05, **p < 0.01, ***p < 0.001

Table 4. Hypotheses Tests and Robustness Checks in Study 2

Variable	Model 1: Baseline model	Model 2: Model 1 + Main effects	Model 3: Model 2 + Moderation effects	Model 4: Objective indicator model	Model 5: Tobit model
Prior exposure	0.014 (0.023)	0.028 (0.022)	0.012 (0.015)	0.012 (0.020)	0.011 (0.013)
Interestingness	-0.014* (0.008)	0.003 (0.008)	0.003 (0.006)	0.001 (0.007)	0.003 (0.005)
Relevance	-0.014 (0.008)	0.001 (0.008)	0.009 (0.006)	0.003 (0.007)	0.006 (0.005)
Pretest believability	0.513*** (0.009)	0.082* (0.043)	0.116*** (0.031)	0.100*** (0.038)	0.099*** (0.028)
Attitude strength	-	0.121*** (0.012)	0.053*** (0.010)	0.107*** (0.011)	0.057*** (0.009)
Claim-source credibility	-	0.023 (0.016)	0.077*** (0.024)	-	0.114*** (0.023)
Fact-checker credibility	-	-0.116*** (0.010)	-0.053** (0.023)	-	-0.034* (0.021)
Fact-check strength (Strong)	-	-0.052** (0.021)	0.030** (0.015)	-0.051*** (0.019)	-
Claim-source credibility × FC credibility	-	-	-0.024** (0.010)	-	-0.033** (0.009)
Attitude strength × Claim-source credibility	-	-	0.062*** (0.002)	-	0.065*** (0.002)
Attitude strength × FC credibility	-	-	-0.038*** (0.002)	-	-0.039*** (0.001)
Reuters (vs. BuzzFeed)	-	-	-	0.009 (0.024)	-
FactCheck (vs. Health Feedback)	-	-	-	-0.158*** (0.024)	-
Reuters × FactCheck	-	-	-	0.006 (0.034)	-
Attitude strength × Reuters	-	-	-	0.063*** (0.004)	-
Attitude strength × FactCheck	-	-	-	-0.043*** (0.004)	-
Intercept	1.171*** (0.233)	2.542*** (0.242)	2.296*** (0.180)	2.352*** (0.223)	2.448*** (0.091)
Log(scale)	-	-	-	-	-1.200*** (0.016)
Observations	2,108	2,108	2,108	2,108	-
R-squared	0.764	0.797	0.897	0.829	-
Adjusted *R*-squared	0.683	0.726	0.861	0.769	-
F-statistic	9.449*** (*df* = 538;1569)	11.326*** (*df* = 542; 1565)	24.941*** (*df* = 545; 1562)	13.876*** (*df* = 545; 1562)	-
Log-likelihood	-	-	-	-	-533.2 (*df* = 12)
Wald-statistic	-	-	-	-	9,702 (*df* = 10)

Note: Standard errors are in parentheses. Fixed effects of participants, claims, and claim order not shown to conserve space. +p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001

Discussion and Conclusions

Key Findings

The goal of this study was to explain the inconsistent effects of corrective fact-checking on user beliefs reported in prior fact-checking research. To explain this inconsistency, I identified three contingent factors from attitude research: claim-source credibility, fact-checker credibility, and attitude strength. The main and moderating effects of these factors were tested using two online experiments on COVID-related misinformation, conducted during earlier and later phases of the pandemic.

Descriptive statistics from the two studies reveal some troubling patterns. First, in both studies, participants viewed claim sources and fact-checkers as less credible, including well-known, reputable organizations like Reuters and FactCheck. This credibility dropped from Study 1 to Study 2, as the COVID-19 pandemic progressed. Moreover, fact-checkers were seen as less credible than claim sources. This suggests that most corrective fact-checks may be dismissed by users as not believable, even if they are true. Under such circumstances, fact-checking may not have the intended effect of correcting false perceptions.

In Study 1, the credibility gap between the well-known, highly acclaimed fact-checker FactCheck and the lesser-known HoaxSlayer was quite slim. This narrow gap suggests that FactCheck's numerous media awards and accolades have not significantly helped to build public credibility. In Study 2, the lesser-known fact-checker, Health Feedback, enjoyed more credibility than FactCheck, which suggests that employing domain experts and having a transparent fact-checking process may be superior to media recognition in helping to build fact-checker credibility.

Notwithstanding the above dismal view of fact-checker credibility, the studies provide empirical support for the role of three contingent factors in influencing users' believability perceptions of fact-checked claims. In both studies, fact-checker credibility significantly influenced users' believability perceptions of online misinformation, individually or jointly with claim-source credibility and attitude strength in the expected direction.

Given that this study examined fact-checking in a public health context, the extent to which its findings may generalize to other contexts (e.g., political) remains unknown. In political contexts, stronger political polarization may lead to a more dominant confirmation bias, where attitude strength may assume greater dominance over claim-source and fact-checker credibility than that observed in this study.

Limitations of the Study

The results of this study should be interpreted in light of its limitations. First, the experimental design employed a simple textual interface rather than colorful graphical online web or social media pages. The concurrent use of text and graphics may confound participant responses if these responses are influenced by the attractiveness of the interface rather than the content of a claim. The simple text-based interface was not subject to such confounding effects.

Second, repeated measures designs are often not preferred over independent group designs because of order effects, where participants' responses to later claims may be conditioned by their exposure to previous claims or by learning or task fatigue. This study controlled for order effects in two ways: (1) by randomizing the ordering of claims and treatments (news sources and fact-checkers) in the experimental design and (2) by partialing out the effects of claims and their ordering using fixed effects models.

Lastly, many other moderators may also influence users' believability perceptions following corrective fact-checks. These moderators may relate to the content of claims, such as their novelty or controversiality, or individual difference variables such as age, education, and online behaviors. I controlled for content-related variance in the analysis by incorporating the interestingness and relevance of claims. While participant demographics, such as age and education were measured, they were excluded from the model because these effects were already factored into the participant-level fixed effects. Nevertheless, I encourage future studies to explore other moderators of fact-checking efficacy.

Theoretical Contributions

The theoretical contributions of this research are threefold. First, this study is one of the earliest to examine the joint effects of all three key parties in the fact-checking process: the claim source, the fact-checker, and fact-check recipients (users). This logic motivated my choice of claim-source credibility, fact-checker credibility, and user attitude strength as the core constructs of interest. While attitude strength has been previously examined in fact-checking research as driving confirmation bias and claim-source credibility has been studied in attitude research, fact-checker credibility is novel to fact-checking research. This study shows that fact-checker credibility helps shape users' beliefs about misinformation, independently and in conjunction with source credibility and users' attitude strength.

Second, this study identifies the pathways by which corrective fact-checking influences user beliefs following corrective fact-checking. The findings suggest that fact-checker credibility impacts claim believability through its interaction with claim-source credibility and attitude strength. If these interactions are taken into account, the main effect of fact-checker credibility may die out. Hence, future fact-checking research should include interaction effects to derive a more nuanced understanding of fact-checking efficacy. Given that there is currently no "theory of fact-checking," the parsimonious set of three predictors (claim-source credibility, fact-checker credibility, and attitude strength) and their interactions may serve as building blocks toward a preliminary theory of fact-checking. Such an approach to theorizing is consistent with the context-specific theorizing advocated by Hong et al. (2014).

Third, this research contributes to the source credibility literature by distinguishing between primary and secondary sources of information (or misinformation) and demonstrating how they may interact, especially when they contradict each other, such as during corrective fact-checking. To the best of my knowledge, this separation of primary and secondary sources is novel to both source credibility and fact-checking literatures and may be extended to non-fact-checking (e.g., advertising) contexts involving multiple sources of conflicting information.

Practical Contributions

This study informs fact-checkers of the reality that most fact-checkers, even those with media recognition or awards, are not viewed in a positive light by online users. Study 2 suggests that staff expertise and process transparency may be more important to building fact-checker credibility than media accolades. Without employing authentic domain experts and transparent fact-checking processes, it is therefore unlikely that fact-checkers will build sufficient credibility to effectively combat misinformation or disinformation.

While there are media rankings and ratings for news sources to guide news consumption behaviors, there are currently no such rankings or ratings for fact-checkers to inform fact-checker credibility perceptions. Since fact-checking credibility influences user beliefs about misinformation, any ranking or rating of fact-checker credibility may help inspire public confidence in at least the more credible fact-checkers.

In conclusion, this study highlights three contingent factors (claim-source credibility, fact-checker credibility, and attitude strength) of relevance to fact-checking of misinformation claims and illustrates how these factors individually and jointly influence user beliefs following corrective fact-checking. Given the nuanced nature of these effects, it would perhaps be wise to reject a simplistic binary view of whether fact-checking works or does not work and instead focus on the underlying conditions that may explain when fact-checking works and when it does not. I also hope that the findings of this study will contribute toward a theory of fact-checking.

References

Amazeen, M. A. (2016). Checking the fact-checkers in 2008: Predicting political ad scrutiny and assessing consistency. *Journal of Political Marketing*, *15*(4), 433-464. https://doi.org/10.1080/15377857.2014.959691

Brannon L. A., Tagler, M. J., & Eagly, A. H. (2007). The moderating role of attitude strength in selective exposure to information. *Journal of Experimental Social Psychology*, *43*(4), 611-617. https://doi.org/10.1016/j.jesp.2006.05.001

Brenan, M. (2021). Americans' trust in media dips to second lowest on record. Gallup. https://news.gallup.com/poll/355526/americans-trust-media-dips-second-lowest-record.aspx.

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

Ecker, U. K, Lewandowsky, S., Cheung, C. S., & Mayberry, M. T. (2015). He did it! She did it! No, she did not! Multiple causal explanations and the continued influence of misinformation. *Journal of Memory and Language*, *85*, 101-115. https://doi.org/10.1016/j.jml.2015.09.002

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.

Fridkin, K., Kenney, P. J., & Wintersieck, A. (2015). Liar, liar, pants on fire: How fact checking influences citizens' reactions to negative advertising. *Political Communication*, *32*(1), 127-151. https://doi.org/10.1080/10584609.2014.914613

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

Grant, M. (1995). *Greek and Roman historians: Information and misinformation*. Routledge. https://doi.org/10.1111/jcom.12038

Hernon, P. (1995). Disinformation and misinformation through the internet: Findings of an exploratory study. *Government Information Quarterly*, *12*(2), 133-139. https://doi.org/10.1016/0740-624X(95)90052-7

Hong, W., Chan, F. K. Y., Thong, J. Y. L., Chasalow, L. C., & Dhillon, G. (2014). A framework and guidelines for context-specific theorizing in information systems research. *Information Systems Research*, *25*(1), 111-136. https://doi.org/10.1287/isre.2013.0501

Hovland, C., & Weiss, W. (1951). The influence of source credibility in communication effectiveness. *Public Opinion Quarterly*, *15*(4), 635-650. https://doi.org/10.1086/266350

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

Howe, L. C., & Krosnick, J. A. (2017). Attitude strength. *Annual Reviews of Psychology*, *68*, 327-351. https://doi.org/10.1146/annurev-psych-122414-033600

Jarman, J. W. (2016). Influence of political affiliation and criticism on the effectiveness of political fact checking. *Communication Research Reports*, *33*(1), 9-15. https://doi.org/10.1080/08824096.2015.1117436

Kearney, M. W. (2017). *Trusting news project report.* Reynolds Journalism Institute. https://www.rjionline.org/reporthtml.html

Kim, A., & Dennis, A. R. (2020). Says who? The effects of presentation format and source rating on fake news in social media. *MIS Quarterly*, *43*(3), 1025-1039. https://dl.acm.org/doi/10.25300/MISQ/2019/15188

Krosnick, J. A., & Abelson, R. P. (1992). The case of measuring attitude strength in surveys. In J. M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (pp. 177-203). Russell Sage Foundation.

Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R. E. Petty & J. A. Krosnick (Ed.), *Attitude strength: Antecedents and consequences*. Erlbaum.

McGinnies, E. (1973). Initial attitude, source credibility, and involvement as factors in persuasion. *Journal of Experimental Social Psychology*, *9*(4), 285-296. https://doi.org/10.1016/0022-1031(73)90066-8

Moravec, P., Kim, A., & Dennis, A. R. (2020). Appealing to sense and sensibility: System 1 and System 2 interventions for fake news on social media. *Information Systems Research*, *31*(3), 987-1006. https://doi.org/10.1287/isre.2020.0927

Moravec, P. L., Minas, R. K., & Dennis, A. R. (2019). Fake news on social media: People believe what they want to believe when it makes no sense at all. *MIS Quarterly*, *43*(4), 1343-1360. https://doi.org/10.25300/MISQ/2019/15505

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175-220. https://doi.org/10.1037/1089-2680.2.2.175

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

Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *PNAS*, *116*(7), 2521-2526. https://doi.org/10.1073/pnas.180678111

Pennycook, G., Bear A., Collins, E. T., & Rand, D. G. (2020). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science*, *66*(21), 4944.4957. https://doi.org/10.1287/mnsc.2019.3478

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes of attitude change*. Springer.

Pingree, R. J., Brossard, D., & McLeod, D. M. (2014). Effects of journalistic adjudication on factual beliefs, news evaluations, information seeking, and epistemic political efficacy. *Mass Communication and Society*, *17*(5), 615-638. https://doi.org/10.1080/15205436.2013.821491

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, *34*(2), 243-281. https://doi.org/10.1111/j.1559-1816.2004.tb02547.x

Porter, E., Wood, T. J., & Kirby, D. (2018). Sex trafficking, Russian infiltration, birth certificates, and pedophilia: A survey experiment correcting fake news. *Journal of Experimental Political Science*, *5*(2), 159-164. https://doi.org/10.1017/XPS.2017.32

Simpson, S. (2019). *Fake news: A global epidemic vast majority (86%) of online global citizens have been exposed to it.* Ipsos Public Affairs. https://www.ipsos.com/en-us/news-polls/cigi-fake-news-global-epidemic

Stencel, M. (2019). Number of fact-checking outlets surges to 188 in more than 60 countries. The Poynter Institute. https://www.poynter.org/fact-checking/2019/number-of-fact-checking-outlets-surges-to-188-in-more-than-60-countries/

Thorson, E. (2016). Belief echoes: The persistent effects of corrected misinformation. *Political Communication*, *33*(3), 460-480. https://doi.org/10.1080/10584609.2015.1102187

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124

Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, *85*(3), 423-441. https://doi.org/10.1080/03637751.2018.1467564

Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2019). Fact-checking: A meta-analysis of what works and for whom. *Political Communication*, *37*(3), 350-375. https://doi.org/10.1080/10584609.2019.1668894

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

Weeks, B. E., & Garrett, R. K. (2014). Electoral consequences of political rumors: Motivated reasoning, candidate rumors, and vote choice during the U.S. presidential election. *International Journal of Public Opinion Research*, *26*(4), 401-422. https://doi.org/10.1093/ijpor/edu005

Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, *32*(2), 303-330. https://doi.org/10.1007/s11109-018-9443-y

Young, D. G., Jamieson, K. H., Poulsen, S., & Goldring, A. (2017). Fact checking effectiveness as a function of format and tone: Evaluating FactCheck.org and FlackCheck.org. *Journalism & Mass Communication Quarterly*, *95*(1), 49-75. https://doi.org/10.1177/1077699017710453

About the Authors

Anol Bhattacherjee is a professor of business analytics and information systems, the Exide Professor of Business Ethics, and the academic director of the Master's of Science Program in Business Analytics and Information Systems at the Muma College of Business, University of South Florida. In 2019, he was awarded the Stafford Beer Medal by the Operational Research Society, U.K. and named the Fulbright-Nehru Distinguished Chair by the United States Department of State. He previously served on the editorial boards of *MIS Quarterly* and the *Journal of the Association of Information Systems*.

Appendix

Measurement Items

Believability (pre-treatment and post-treatment):

- (Pretreatment) To the best of your knowledge, …
- (Post-treatment) Considering the claim source and fact-check rating, …
- How truthful do you find the above claim? (mostly false to mostly true)
- How convincing do you find the above claim? (extremely unconvincing to extremely convincing)
- How believable do you find the above claim? (not believable at all to extremely believable)

Credibility of news sources and fact-checkers:

- To what extent do you believe that the following websites are trustworthy? (mostly untrustworthy to mostly trustworthy)
- To what extent do you believe that the following websites have the necessary expertise to do their job? (no expertise at all to high level of expertise)
- To what extent do you find the following websites credible? (mostly non-credible to mostly credible)

Attitude strength:

- How important do you find the above claim? (not important at all to extremely important)
- How strongly do you feel about the above claim? (not at all strongly to extremely strongly)
- How certain are you about your attitude toward the above claim (highly uncertain to highly certain)

- **Control Variables:**

- Prior exposure: Please indicate if you have seen the above claim in the media before this study (no | yes)
- Interestingness: How interesting do you find the above claim? (not interesting at all to very interesting)
- Relevance: To what extent is the above claim relevant to people in your social network? (not relevant at all to extremely relevant)

COVID-19 Claims

Study 1:

1. Contact tracing can reduce the spread of COVID-19.
2. CDC recommends wearing two masks for adequate protection against COVID-19.
3. COVID-19 virus has been detected in human feces and wastewater.
4. A cheap, widely available drug called dexamethasone provides effective COVID-19 relief among severely sick patients.
5. Contact lens disinfecting solution can kill the COVID-19 virus on contact.
6. Vaccines may not prevent people from contracting COVID-19—Dr. Fauci.
7. People who have recovered from COVID-19 have acquired immunity to the disease.
8. Doctors say that children are at lower risk of contracting COVID-19 than adults.
9. WHO: People with autoimmune and other serious diseases should avoid COVID-19 vaccines for now.
10. The drug remdesivir is known to reduce deaths among COVID-19 patients.

Study 2:

1. COVID-19 vaccines provide stronger immunity to reinfection than prior infections.
2. If you had a previous COVID-19 infection, the CDC does not recommend that you take a vaccine.
3. If you took both COVID-19 vaccines and the booster vaccine, you are safe against COVID-19.
4. Vaccination reduces the risks of long-term COVID-19.
5. At-home rapid COVID-19 testing kits are fairly accurate.
6. Cannabis helps reduce the chances of COVID-19 infection.