

An experimental study of the effectiveness of fact checks

Tsang, Stephanie Jean; Zheng, Jingwei; Li, Wenshu; Salaudeen, Mistura Adebusola

Published in:
Online Information Review

DOI:
[10.1108/OIR-09-2022-0492](https://doi.org/10.1108/OIR-09-2022-0492)

Published: 08/11/2023

Document Version:
Peer reviewed version

[Link to publication](#)

Citation for published version (APA):
Tsang, S. J., Zheng, J., Li, W., & Salaudeen, M. A. (2023). An experimental study of the effectiveness of fact checks: interplay of evidence type, veracity and news agreement. *Online Information Review*, 47(7), 1415-1429. <https://doi.org/10.1108/OIR-09-2022-0492>

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent publication URLs



**An Experimental Study of the Effectiveness of Fact Checks:
Interplay of Evidence Type, Veracity, and News Agreement**

Journal:	<i>Online Information Review</i>
Manuscript ID	OIR-09-2022-0492.R1
Manuscript Type:	Research Paper
Keywords:	evidence type, fact-checking, psychological reactance, political communication, veracity

SCHOLARONE™
Manuscripts

Abstract

Purpose – Given the rapid growth in efforts on misinformation correction, the study aims to test how evidence type and veracity interact with news agreement on the effectiveness of fact-checking on how well a corrective message discount a false news information.

Methodology – Experimental participants ($N = 511$) in Hong Kong were exposed to the same news article and then to a piece of corrective information debunking the news article with variation in the types of evidence (numerical vs. narrative) and veracity (no verdict vs. half false vs. entirely false) in 2019.

Findings – Among the participants who disagreed with the news article, numerical fact-checking was more effective than narrative fact-checking in discounting the news article. Some evidence of the backfire effect was found among participants for whom the article was attitude-incongruent.

Originality – When debunking false information with people exposed to attitude-incongruent news, a milder verdict presented in the form of a half-false scale can prompt a more positive perception of the issue at stake than an entirely false scale, implying that a less certain verdict can help in mitigating the backfire effect compared to a certain verdict.

Keywords: evidence type, fact-checking, psychological reactance, political communication, veracity, Hong Kong

An Experimental Study of the Effectiveness of Fact Checks:
Interplay of Evidence Type, Veracity, and News Agreement

To contain and limit the impact of misinformation, fact-checking efforts have been active in debunking false information to mitigate the impact of falsehoods spread online (Bell, 2019; Graves and Cherubini, 2016). Although the last 10 years have marked rapid growth in research on misinformation correction, ranging from institutional fact-checking practices to individual reception of corrective messages (e.g., Graves, 2018; Graves et al., 2016; Kim et al., 2022; Shin and Thorson, 2017), the effects of misinformation correction are not yet thoroughly understood. Walter et al.'s (2020) meta-analysis yielded mixed results on the effectiveness of fact checks in correcting misperceptions, and the authors called for taking more individual-level factors into account, including political interest and political efficacy. While prior studies have investigated the effects of different message factors of corrective information, such as the presence of a truth scale, lexical complexity, and message length (Amazeen et al., 2018; Nyhan and Reifler, 2012; see also Walter et al., 2020), this study aimed to contribute to this growing body of literature by testing the effectiveness of both message factors (evidence type and veracity) and individual-level factors (level of news agreement).

Since fact-checking involves “providing analyses and assessments of claims and content” to the public (Brandtzaeg et al., 2018, p. 1111), the delivery of assessments is vital for correcting misperceptions. Indeed, the presentation of fact-checking has been shown to impact the attention a message receives, message discounting, and argument scrutiny (Young et al., 2018). Therefore, this study extended this line of research by investigating whether the use of statistics and testimonies would have different effects on people's willingness to discount a news article as false. In addition, we suspected that displaying a definite verdict (i.e., entirely false) would likely

cause backfiring, sometimes also called the boomerang effect (Miller, 2015), whereas this would be less likely with a less definite verdict (i.e., half false). In other words, telling readers that a piece of news is half false or entirely false might have different outcomes regarding their perceptions of the news being debunked as well as their evaluations toward the incident discussed in the news article. The findings of this study will shed light on the conditions that improve the effectiveness of misinformation correction in a political context.

To explore the effects of evidence type and veracity in a piece of corrective information on audiences' perceptions of a news message about a public policy and evaluations of that policy, we conducted an online experimental survey. The survey employed a 2 (evidence type: numerical vs. narrative) \times 3 (veracity: no verdict (control) vs. half false vs. entirely false) \times 3 (news agreement: agree vs. neither agree nor disagree (control) vs. disagree) experimental design with a sample of Hong Kong adults ($N = 511$). The results indicate that among participants who initially disagreed with the news article on the negative influence of immigrants, fact checks debunking the exaggerated negative influence caused with numerical evidence were more effective than the same fact checks with narrative evidence in discounting the false news message. In addition, participants for whom the news article was attitude-incongruent were more supportive of the pro-immigrant policy when they were exposed to a half-false verdict than to an entirely false verdict. The paper will discuss the implications for advancement of the delivery of fact checks to correct public misperceptions.

Numerical vs. Narrative Evidence

Since fact-checking is a form of evidence-based persuasion to correct public misperceptions (Garrett and Weeks, 2013), it makes sense to ponder whether certain types of evidence make fact-checking more efficient at debunking false information than others. For

instance, Young et al. (2018) found that fact checks presented as videos with narration were more efficient than print for changing readers' beliefs, and the inclusion of humor did not play a role in influencing policy evaluations. Chen et al. (2022) found that debunking messages with more authoritative sources were able to lower the possibility of information distortion by ending a discussion thread quicker on WhatsApp. Moreover, Fridkin et al. (2015) found that negative fact checks disputing the claims of a commercial were more robust than positive fact checks agreeing with the claims. In the present study, we focused on the evidence provided for fact-checking to test whether numerical debunking evidence (i.e., statistics) would be more powerful than narrative debunking evidence (i.e., testimonies).

Comparisons of numerical and narrative evidence are common in the persuasion literature (Hong and Park, 2012). Often, the forms of evidence can impact how audiences evaluate the credibility of news (Henke et al., 2020). Numerical evidence refers to a quantitative description that “provides proof in a form of summary information across a larger number of cases” (Allen and Preiss, 1997, p. 126) or “information on an object, person, concept or phenomena of a statistical (base rate) or numerical nature” (Church and Wilbanks, 1986, as cited in Kazoleas, 1993, p. 40). Narrative evidence, however, refers to materials like testimonials, anecdotes, and examples (Kazoleas, 1993). Some meta-analyses have suggested that numerical evidence is more persuasive (Allen and Preiss, 1997; Gray and Harrington, 2010), while other research has shown that narrative evidence may be more persuasive (Kreuter et al., 2010; Murphy et al., 2013).

Compared to narrative evidence, numerical evidence is regarded as more credible, as statistical facts are based on rationality (Baesler and Burgoon, 1994). Overall, statistical evidence based on rationality might be more successful at revealing a news article to be a piece of false

information. Young et al. (2018) constructed message discounting using Nabi et al.'s (2007) scale to assess the extent to which people discounted corrective stimuli as “just a joke” (p. 59). In the current study, we assessed the degree to which people discounted a news stimuli by identifying the news as “fake,” and we expected that when asked about whether the news was fabricated, individuals exposed to a numerical fact check would discount the news to a larger extent than those exposed to a narrative fact check:

H1: Compared to participants exposed to narrative evidence, participants exposed to numerical evidence will discount the news article to a greater extent.

Furthermore, we anticipated that the use of statistics would have a stronger impact on policy evaluations than qualitative data would (i.e., narrative descriptions of facts). Individuals tend to assume that numerical evidence is unbiased and impersonal proof (Allen and Preiss, 1997). Strongly objective information has been found to have greater perceived value (Green and Brinn, 2003) than narrative evidence. Hence, we expected numerical fact checks to be more powerful in facilitating higher policy evaluations and that participants exposed to a numerical fact check would perceive the policy discussed in the news stimulus as more favorable than participants exposed to a narrative fact check:

H2: Compared to participants exposed to narrative evidence, participants exposed to numerical evidence will evaluate the policy positively to a larger extent.

Truth Scales and Veracity

Given that the majority of fact-checking reports in the market assess the veracity of public claims (Graves et al., 2016) and conclude whether the claim being evaluated is true or false with a verdict (see Mena, 2019; e.g., true, half false, false, etc.), veracity was also considered in this study. According to Miller (2015), messages aimed at controlling people's

thoughts and behaviors are prone to arousing resistance. Stemming from the theory of psychological reactance (Brehm and Brehm, 1981), individuals believe they possess certain freedoms to exhibit free behaviors. Then, when they perceive a threat to that freedom, an unpleasant motivational arousal (reactance) emerges (Brehm and Brehm, 1981, 2013). More importantly, language can lead media audiences to feel that their freedom is under threat, arousing anger and negative cognition regarding the message (Dillard and Shen, 2005).

Elaborating on this idea, we suggest that a corrective message can be perceived as a threat to the individual's psychological freedom to evaluate the truthfulness of news. When individuals receive fact-checking alerts that a piece of news is false and/or a belief is incorrect, they might see this as asking them to adopt or abandon a position (Brehm and Brehm, 1981). Consequently, they will likely show reactance—that is, do something to re-establish their freedom. According to the persuasion literature, reactance might lead to negative attitudes about a persuading message and less willingness to behave accordingly and, sometimes, even displays of anger and counterarguments against the message (Kim et al., 2013).

Since a verdict about a news message often comes in the form of a truth scale, such as the one launched by PolitiFact (2018)—true, mostly true, half true, mostly false, false, and pants on fire—as well as the one developed by Snopes (2021)—true, mostly true, mixture, mostly false, false, unproven, etc.—this study examined veracity with visuals. A truth scale is a form of rating scale with “a clear visual indicator” to help readers quickly ascertain the relative accuracy of a claim being fact-checked (Amazeen et al., 2018). According to Stencel (2016), most fact-checking outlets in the United States employ some sort of truth meter. The simple, straightforward verdicts a rating scale offers allow readers to quickly determine whether a given

statement or news article is true or false. Previous studies have found that visual truth scales make fact checks more effective, particularly in non-political contexts (Amazeen et al., 2018).

In Asia, similar truth scales are used by professional fact-checkers. Alongside other labels such as “no evidence,” Taiwan FactCheck Center (2020) present their verdicts using true, partly false, and false. In Hong Kong, Factcheck Lab (2022) and HKBU Fact Check (2022) use true, partly false, and false, alongside “misleading” and “no evidence.” Our study therefore employed truth scales as a tool for signaling veracity, either as entirely false or half false, with no presentation of a verdict as a control. We expected the presence of an entirely false verdict to make the correction less effective when a verdict was uncertain. When reactance is aroused, people are motivated to restore lost freedom either by derogating the source of the threat (Rains, 2013) or by comprehending the threatened choice as more attractive (Brehm and Brehm, 1981). Further, Borah et al. (2022) found that the arousal of negative emotions towards corrective messages could lead to higher levels of COVID-19 misperceptions. Since a more extreme verdict, “entirely false,” indicates the original message is false to a larger degree, it is anticipated to stimulate more negative reactions than a less extreme verdict, “half false.” By shedding light on veracity via the presence of truth scales, we expected our findings not only to inform the literature on backfire effects but also to elucidate the effectiveness of the different veracity ratings in practice:

H3: Compared to participants exposed to a half-false verdict and no verdict, participants exposed to an entirely false will discount the news article to a lesser extent.

H4: Compared to participants exposed to an entirely false verdict, participants exposed to a half-false verdict and no verdict will evaluate the policy positively to a larger extent.

News Agreement and Confirmation Bias

When it comes to perceiving information, in the contexts of both the processing and reception of a news article or corrective information, the reader's ideology must be considered (Porter, Wood, and Kirby, 2018), as it can impact their processing of information. According to Clyton et al. (2019), people with different party identifications tend to hold distinct perceptions of the truthfulness of news on the same issue, and this, in turn, prompts different reactions to fact-checking. This can be explained by confirmation bias (Mena, 2019), according to which people tend to process information in a biased manner. That is, people tend to accept attitude-congruent, reinforcing information at face value and tend to reject information that counters their personal stance, even when it presents concrete evidence. Consistent with findings on hostile media perception (Vallone et al., 1985), in which audiences holding opposing points of view were found to evaluate the exact same news message as fake to significantly different degrees (Tsang, 2021; 2022b). In this sense, whether a fact check can discount a news message and correct beliefs largely depends on the readers' prior beliefs. As Mena (2019) suggested, it is important to explore the implications of this for fact-checking effectiveness.

Thus, confirmation bias and hostile media perception explain why fact-checking may sometimes backfire and politically interested partisans may experience less change in their misinformed beliefs (Nyhan and Reifler, 2010). The term "backfiring" refers to a phenomenon in which the message receiver's belief moves in the direction opposite to the position advocated in a message such as a health warning label, an environmental message, and a media literacy intervention (Byrne and Hart, 2009). Whether people accept a fact check and change their beliefs often depends on their drive to protect their preexisting beliefs and reaction to threats to their self-identity. Although research has found that backfire effects can be mixed (see Lewandowsky et al., 2020; Wood and Porter, 2019), we expected news agreement (i.e., whether one's prior

attitude contrasts with the position advocated in a news article) to have a negative influence on whether people discount the news article after it has been debunked with corrective information and subsequently change their beliefs. The hypotheses below follow the long line of literature on confirmation bias and hostile media perception as well as the call to investigate the interplay of the display of fact-checking, message beliefs, and credibility judgments (Walter and Murphy, 2018):

H5: There will be a negative main effect of news agreement on message discounting.

H6: There will be a negative main effect of news agreement on policy evaluations.

Three-Way Interactions

Corrective information with a truth scale has been shown to be effective in certain circumstances. Thus, examining the interplay of evidence type and verdict presentation is critical for developing better fact-check designs. For instance, evidence type and veracity might interact differently for people who disagree with a news article than for those who agree with the news article. Individuals motivated not to protect their belief system but to arrive at an accurate conclusion might crave numerical evidence, which they see as more persuasive in general (Gray and Harrington, 2010). In contrast, those motivated to reinforce their prior beliefs might be more attracted to narrative evidence, which is more vivid (Zillmann and Brosius, 2000). Since there are a few possible outcomes of news agreement (i.e., reinforcement of previous beliefs or backfire effects), we formulated the following research question:

RQ1: How does news agreement impact the interplay of evidence type and veracity regarding message discounting?

The presence of a visual truth scale can make the verdict and information in the fact check easier to process and understand (Amazeen et al., 2018). Indeed, as supported by the

literature on persuasion (Cacioppo and Petty, 1982; Petty and Cacioppo, 1986), presenting an audience with graphical content encourages them to take the “periphery route” of cognitive processing, which decreases their motivation to scrutinize the details of a claim and instead rely solely on the graphical information (Nyhan and Reifler, 2011). However, those who agree with the news article might not read the information in a counter-attitudinal fact check deeply, potentially leading them to immediately reject the correction and adhere to the belief that the news article is legitimate. Therefore, a different presentation of a verdict may result in backfire effects for fact-checking organizations when the corrective messages contradict the audience’s partisan disposition. As stated previously, half-false and entirely false verdicts can trigger different levels of psychological reactance and backfiring. The fact that the absence of a verdict, a half-false verdict, and an entirely false verdict might interact with the other two variables in varying manners led to the second research question:

RQ2: How does news agreement impact the interplay of evidence type and veracity on policy evaluations?

Method

Sample and Procedure

To achieve the research objectives, an online experiment was conducted in Hong Kong in March 2019. Dynata (previously known as SSI) was responsible for data collection, and participants were paid for their participation. To ensure that the focal immigration policy would be of similar relevance to participants, only Hong Kong citizens were invited to participate. Participants were first asked to sign a consent form to ensure voluntary participation. They were then asked to read a news article on an immigration controversy and answer manipulation check questions regarding the topic of the news article. Next, participants were asked to read a

corrective message and answer some questions about both this corrective message and the news article. Finally, participants were debriefed. The debriefing process clarified that the messages they had seen were manipulated for the purpose of the study. Since 16 participants failed to complete the entire experiment, the final sample included 511 participants, 263 of whom were female (51.5%). The average age range was 38–47 years old. The majority of participants were college graduates, and the average monthly family income was HKD 20,000–29,999.

Design and Stimulus

The experiment used a 3 (veracity: no verdict vs. half-false verdict vs. entirely false verdict) \times 2 (evidence: numerical vs. narrative) \times 3 (measured news agreement: news disagreement vs. neutral vs. news agreement) design. After reading a news article on immigration policy, participants were prompted to assess the degree to which they agreed with the news article on a scale from 1 to 5, with 1 being “completely disagree” and 5 being “completely agree.” To measure news agreement, the sample ($M = 3.66$, $SD = .90$) was split into three groups for comparison purposes. The first group contained 45 participants (8.8%) who completely or slightly agreed with the news article. The second group comprised 307 participants (60%) who completely or slightly disagreed with the corrective message. The third group, which was the control group (neutral), contained 160 participants (31.3%) who neither agreed nor disagreed with the corrective message.

The fabricated news article featured an exaggeration of the estimated number of mainland Chinese immigrants who have moved to Hong Kong over the past 20 years. We intentionally doubled the number of immigrants originally published in *Apple Daily* in 2018 and exaggerated the negative influence of immigrants, replicating most of the disinformation published in Hong Kong on this issue. More specifically, the news article stated, “If there were no ‘one-way’ permit

scheme for migrants to move from mainland China into Hong Kong, Hong Kong would have had 1.5 million fewer workers over the past 20 years.” The People’s Republic of China Permit for Proceeding to Hong Kong and Macao (commonly known as a one-way permit) is issued by the Public Security Bureau in mainland China (Immigration Department, 2015) and is mainly used for family reunification; mainland–Hong Kong cross-boundary marriages account for about one third of the registered marriages in Hong Kong (Census and Statistics Department, 2018). The current daily quota for such permits is 150. Since 1997, over a million mainland residents have settled in Hong Kong through this route (The Government of the Hong Kong Special Administrative Region, 2019). Some locals have argued that the high number of mainland residents settling in Hong Kong has imposed a heavy burden on Hong Kong’s public services, such as housing and healthcare.

All participants were randomly assigned to one of six manipulated conditions. Eighty-four participants (16.4%) read a numerical fact check with no verdict (control condition), 84 participants (16.4%) read a numerical fact check with a half-false verdict, and 85 participants (16.6%) read a numerical fact check with an entirely false verdict. The remaining participants read a narrative fact check with no verdict ($n = 86$, 16.8%), a half-false verdict ($n = 85$, 16.6%), or an entirely false verdict ($n = 88$, 17.2%). The content, length, and layout of the fact check remained consistent across conditions to ensure internal validity. Participants were debriefed message at the end of experiment. The message informed all participants the news report seen was fabricated, and participants should seek official documents to be informed about the immigration status in Hong Kong.

The fact check featured the headline “[Fact Check] 1.5 million migrants moved into Hong Kong due to the ‘one-way’ permit scheme since the turnover to China?” Information was

given in either numerical or narrative form to debunk whether 1.5 million new immigrants had flooded into Hong Kong in the past 20 years and whether the one-way permit scheme has resulted in a major reason for the Hong Kong housing crisis. The numerical evidence relied solely on statistical reports from three authoritative sources, whereas the narrative fact check took the form of anecdotal or personal evidence, such as interviews, exemplars, stories, testimonials, and opinions (Han and Fink, 2012; Tversky and Kahneman, 1974). Instead of three sets of statistics, the narrative fact checks utilized three testimonials to debunk the news article.

Further, the corrective information was accompanied by a half-false verdict, an entirely false verdict, or no verdict. The presentation of the verdict adapted truth scales found on prominent fact-checking sites to fit the Chinese language context. The verdict was presented with a pointer and a panel with two directions: true and false. The pointer of the half-false verdict pointed to the mid-point between true and false.

Measures

Message discounting. We examined the fact-checking effect by measuring the extent of the audience's message discounting, operationalized as perceived fakeness of the news (Tsang, 2021). As the fact-check information advised participants about the veracity of the news article, we asked them to report the extent to which they thought (a) the news contained misleading components, (b) the news exaggerated the negative consequences of new immigrants brought by the policy (i.e., number of immigrants entering Hong Kong), (c) the news contained critical errors, (d) the presented facts (i.e., numbers) were fictitious, and (e) the content was fabricated. The five items were rated on a five-point Likert scale, with 1 being "completely disagree" and 5 being "completely agree" ($M = 2.14$, $SD = .89$, Cronbach's $\alpha = .88$). Higher numbers signal

higher degrees of perceived fakeness of news, which equals to higher degrees of message discounting.

Policy evaluation. We assessed the effect of policy evaluation by measuring audience support for the one-way permit policy. Since the news fabricated data about the negative aspects of the one-way permit policy, a successful debunk of the news should lead to a better evaluation of the policy, and vice versa. Using a seven-point scale, participants were asked whether the policy was harmful (1) or beneficial (7) and whether the policy was valueless (1) or valuable (7). The mean of the two items was used as a measure of policy evaluation ($M = 2.93$, $SD = 1.48$, Cronbach's $\alpha = .94$). Higher numbers signal more positive evaluations toward the policy.

Political interest. Participants were asked to self-report the extent to which they were interested in politics and current affairs using a five-point Likert scale, with 1 being “completely disagree” and 5 being “completely agree” ($M = 2.62$, $SD = 1.01$).

Party identification. Lastly, participants were asked to indicate whether they were pro-democracy ($n = 140$, 27.4%) or pro-establishment ($n = 54$, 10.6%), agreed with both sides, or were unsure. Those who indicated agreement with both sides and those who were unsure were combined into one “neutral” category ($n = 317$, 62.0%).

Results

To explore the interaction effects of evidence type, veracity, and news agreement on message discounting (RQ1) and policy evaluation (RQ2), two three-way analyses of covariance (ANCOVAs) were run with covariates (party identification and political interest). Covariates were included because news agreement was measured rather than manipulated and random assignment was not possible. A series of pairwise comparisons with Bonferroni corrections was run to test H1, H3, and H5 concerning message discounting and H2, H4, and H6 concerning

policy evaluation with three independent groups (people who disagree with the news, people who are neutral, and people who agree with the news).

Message Discounting

To test the three hypotheses regarding message discounting (H1, H3, H5), an ANCOVA was run that included party identification ($F(1, 491) = 1.29, p = .257$) and political interest ($F(1, 491) = 2.76, p = .097$) as covariates. As no three-way interaction ($F(2, 491) = .355, p = .841$) was found and none of the three two-way interactions were significant at the .05 level, only main effects are reported here. In general, when given a fact check after news exposure, participants presented with numerical evidence ($M = 2.21, SD = .93$) were more likely than participants presented with narrative evidence ($M = 2.08, SD = .85; F(1, 491) = 6.66, p = .010, \text{partial eta squared} = .013$) to rate the news story as fake. Therefore, H1 was supported; numerical evidence made a piece of corrective information more effective than narrative evidence did with respect to message discounting.

As expected, news agreement had a main effect on message discounting; those who disagreed with the news viewed it as the most fake ($M = 2.91, SD = .90$), while those who agreed with the news viewed it as the least fake ($M = 1.91, SD = .86$). Neutral participants' scores were between those of the other two groups ($M = 2.37, SD = .79; F(2, 491) = 34.58, p < .001, \text{partial eta squared} = .123$). The three groups of participants were found to differ (disagree vs. neutral at $p = .007$, neutral vs. agree at $p < .001$, disagree vs. agree at $p < .001$). Thus, H5 was supported. Since the truth scale had no main effect ($F(2, 491) = .812, p = .445$), H3 was not supported.

To answer RQ1, the sample was split into groups according to news agreement (agree, neutral, and disagree), and a series of pairwise comparisons adjusted by the Bonferroni method was run within each group. Among participants who disagreed with the news, those exposed to

numerical evidence ($M = 3.34$, $SD = .98$) perceived higher levels of fakeness than those exposed to narrative evidence ($M = 2.28$, $SD = .69$) but only in the entirely false condition ($p = .040$). As shown in Figure 1, such a difference was not found for participants in the control ($p = .904$) and half-false conditions ($p = .901$). Among participants who were neutral, evidence type ($F(1, 150) = .357$, $p = .551$), veracity ($F(2, 150) = .357$, $p = .866$), and interaction ($F(2, 150) = .085$, $p = .919$) did not have any effect on message discounting. The same was true for participants who agreed with the news; there was no interaction effect of evidence type and veracity ($F(2, 296) = .100$, $p = .905$) or main effects of evidence type ($F(df = 1, 296) = 2.29$, $p = .132$) or veracity ($F(2, 296) = 1.44$, $p = .239$) on message discounting.

[Insert Figure 1 about here.]

Policy Evaluation

To test the other three hypotheses (H2, H4, and H6) using policy evaluation as the criterion variable, another set of ANCOVAs was run, controlling for party identification and interest. The results show that the three-way interaction had a significant impact on policy evaluation ($F(4, 491) = 2.50$, $p = .042$). Hence, instead of reporting the two-way interactions and main effects, we moved directly to splitting the file into three groups according to perceived news agreement (disagree, neutral, agree) and looked for two-way interactions within each group (RQ2). Among participants who disagreed with the news article, evidence type was found to interact with veracity ($F(1, 36) = 6.43$, $p = .016$, partial eta squared = .155). Further pairwise comparisons that were adjusted using the Bonferroni method suggested that participants exposed to numerical evidence ($M = 6.00$, $SD = .61$) were more supportive of the policy than participants exposed to narrative evidence ($M = 2.71$, $SD = 1.67$) but only when there was no verdict ($p = .002$; see Figure 2).

[Insert Figure 2 about here.]

Among participants who were neutral, neither evidence type ($F(1, 150) = .003, p = .954$) nor veracity ($F(2, 150) = .330, p = .720$) had any effect on policy evaluation. Lastly, among those who agreed with the news article, truth scale had a main effect on policy evaluation ($F(2, 296) = 3.07, p = .048, \text{partial eta squared} = .020$). Pairwise comparison suggested that participants exposed to the half-false verdict ($M = 3.22, SD = 1.37$) were more supportive of the policy than those exposed to the entirely false verdict ($M = 2.47, SD = 1.34$) in the numerical evidence condition ($p = .031$) but not the narrative condition ($p = 1.000$).

Discussion

Misinformation can influence people's thinking and decision-making even after they have received corrective messages. Hence, debunking information might not be able to change people's beliefs about whether a claim they saw was fake or their evaluations about the policy discussed in the news article. This study involved an online experiment investigating the interplay of news article discrepancy, evidence used in the fact check, and veracity as presented in the form of a truth scale within the context of fact checks about immigration issues in Hong Kong. Overall, there was limited evidence of backfire effects, consistent with Lewandowsky et al. (2020) and Wood and Porter (2019), such that the fact check did not drive stronger misconceptions among those who initially agreed with the news. Research on hostile media perception has argued that motivated reasoning plays a significant role in audiences' perceptions of bias in news (Vallone et al., 1958), and we found the same in relation to the perceived fakeness of a news article. As people tend to use attitudinal-consistent fact checks as references to support their personal points of view (Brandtzaeg et al., 2018), news agreement had a main effect on news message discounting. Those who disagreed with the news article evaluated it as

more fake than those who agreed with it, with neutral participants scoring between the two poles. The pattern was so strong that all the differences between groups were found to be significant.

Our findings suggest that when participants were presented with a fact check, numerical evidence (i.e., statistics) was generally more effective than narrative evidence (i.e., testimonials) for making them believe that what they read was fake news. This is consistent with the literature, which has shown that statistical evidence can be stronger in these cases (Allen and Preiss, 1997; Gray and Harrington, 2011). However, among participants who disagreed with the news article, numerical evidence was significantly more effective than narrative evidence only in the entirely false condition. Since such a difference was not found among participants in the control and half-false conditions, our findings imply that a direct (i.e., more extreme) verdict with numerical evidence was most effective on people who disagreed with the news. In other words, compared to a narrative fact check, numerical evidence more effectively reinforced these people's thought that the news was fake. Evidence type and veracity had no impact on perceived news fakeness among participants who agreed with or were neutral about the news article. Manipulating the evidence type and veracity only impacted those who agreed with the news article. Therefore, prior attitude toward the issue of interest (news agreement) is an important factor to consider when considering the potential impact of different components of any debunking information. Accordingly, we suggest using numerical evidence when trying to convince the public as to what is right or wrong via fact checks.

Regarding policy evaluation or the entity mentioned in a claim, news agreement plays a substantial role. Among participants who disagreed with the news article, numerical evidence drove more policy support than narrative evidence did, especially in the absence of a verdict. In contrast, participants who agreed with the news article might have thought, "Why is the news I

agree with being fact-checked?” The presence of a fact check might have signaled that something was wrong with their beliefs, diminishing their view of the policy. Without the presence of a verdict, the difference between numerical and narrative evidence was greatest, suggesting a similar logic as above: for people receiving an attitude-consistent fact check, statistics were more effective than testimonies. Among participants exposed to numerical evidence, those not exposed to a verdict were more supportive of the policy than participants who were exposed to the entirely false verdict. A post hoc rationale for this finding could be that the presence of a fact check, not just the information provided (news article and fact-checking information), influenced the image of the policy. Among the neutral participants, neither evidence type nor veracity had any effect on belief (i.e., policy evaluation). These results echo Mena’s (2019) finding that confirmation bias can be expected to impact how people process a fact check and, in turn, evaluate the entity mentioned in that particular fact check.

Lastly, among participants who agreed with the news article, those exposed to the half-false verdict were more supportive of the policy than were those exposed to the entirely false verdict. The half-false verdict was particularly effective when participants were presented with numerical evidence. Compared to the half-false verdict, the entirely false verdict tended to drive less support for the policy, indicating mild backfire effects. In sum, the findings shed light on existing theoretical discussions about motivated reasoning in the realm of fact-checking. On the one hand, since a half-false verdict was found to reduce psychological reactance among news audiences, the strategy to reduce extreme reactance concerning veracity can enhance a fact check’s effectiveness on individuals who agree with the piece of misinformation in the first place. On the other hand, tagging corrective messages as entirely false was found to motivate participants to apply more analytical thinking, which can facilitate the generation of more

accurate news evaluations (Pennycook and Rand, 2019). In addition, numerical fact checks generally were more effective than narrative fact checks, given the capability of the former to suppress individuals' psychological reactance. Indeed, fact checks often guide users on the credibility of news content through the supply of detailed analyses of the veracity of a claim (Brandtzaeg et al., 2018). Consider, for example, the truth scale developed by PolitiFact (2018). When a claim is false, fact-checkers can pick from half true, mostly false, false, and pants on fire. While mostly false and false can be used to convey that a claim contains false information, the use of a milder verdict might be helpful in rebuilding the image of the entity mentioned in the debunked claim.

Implications

In sum, compared to narrative evidence, numerical evidence should be provided whenever available to debunk a news message. Regardless of the audience's initial agreement with a news article, the presence of statistics has been shown to exert more prevailing fact-checking effects on message discounting and building more positive attitudes toward the entity involved. Further, when dealing with content that includes both true and false information, fact-checking organizations should consider adopting a milder tone and language to signal falsehood in the verdict (i.e., half-false), rather than an extreme tone (i.e., entirely false). Such a suggestion does not mean that fact-checking organizations should sacrifice accuracy when evaluating the fakeness of corrective message. Given that misinformation and disinformation often involve the use of factual information in a misleading or inaccurate manner (Tsang, 2022a), it would be ideal for fact-checking organizations to acknowledge the truthful components while pointing out the misleading or inaccurate aspects in a courteous manner. Such a strategy could be helpful for mitigating possible effects of driving hostility and psychological reactance. As polarization has

become a major concern in most societies, fact-checkers should continue to develop better strategies to convey corrective messages to politically sensitive audiences. Besides how verdicts are communicated, these findings also shed lights how verdict scales should be developed.

Limitations

Despite the above contributions, this research had its share of shortcomings. Since the data were collected in 2019, the characteristics of Hong Kong media, the fact-checking industry, and public perceptions of fact checks might have changed due to the rise of fact-checking in the city. It should also be noted that the political environment was very polarized during data collection, right before the protests in 2019. Moreover, researchers should continue to conduct similar research in varying contexts and diverse societies, as different cultures and common understandings of fact-checking may affect how individuals interpret corrective messages and truth scales.

Conclusion

In conclusion, this study contributes to research supporting the interaction between evidence type and veracity on fact-checking effectiveness, taking audiences' initial agreement with the piece of misinformation into account. When communicating with individuals who disagree with a news article, numerical fact checks are more effective than narrative corrective information in shaping the perception that the news is indeed false. In addition, when debunking false information with people exposed to attitude-incongruent news, a milder verdict presented in the form of a half-false scale can prompt a more positive perception of the issue at stake than an entirely false scale, implying that a less certain verdict (half false) can help in mitigating the backfire effect compared to a certain verdict (entirely false). With more organizations around the world providing fact-checking services, scholars should continue to advance knowledge to

enhance the effectiveness of misinformation correction, which will likely assist in combatting the spread of misinformation and, in turn, contribute to the functioning of democracies.

References

- Allen, M., Bruflat, R., Fucilla, R., Kramer, M., McKellips, S., Ryan, D. J., and Spiegelhoff, M. (2000), "Testing the persuasiveness of evidence: Combining narrative and statistical forms," *Communication Research Reports*, Vol. 17 No. 4, pp.331-336.
- Allen, M., and Preiss, R. W. (1997), "Comparing the persuasiveness of narrative and statistical evidence using meta-analysis," *Communication Research Reports*, Vol. 14 No. 2, pp.125-131. <https://doi.org/10.1080/08824099709388654>
- Allen, M., Preiss, R. W., and Gayle, B. M. (2006), "Meta-analytic examination of the base-rate fallacy," *Communication Research Reports*, Vol. 23 No. 1, pp.45-51.
- Amazeen, M. A., Thorson, E., Muddiman, A., and Graves, L. (2018), "Correcting political and consumer misperceptions: The effectiveness and effects of rating scale versus contextual correction formats," *Journalism and Mass Communication Quarterly*, Vol. 95 No. 1, pp.28-48. <https://doi.org/10.1177/1077699016678186>
- Baesler, E. J., and Burgoon, J. K. (1994), "The temporal effects of story and statistical evidence on belief change," *Communication Research*, Vol. 21 No. 5, pp.582-602.
<https://doi.org/10.1177/009365094021005002>
- Bell, E. (2019), The fact-check industry," *Columbia Journalism Review*.
https://www.cjr.org/special_report/fact-check-industry-twitter.php
- Borah, P., Kim, S., and Hsu, Y. C. (2022), "'Masks do not work': COVID-19 misperceptions and theory-driven corrective strategies on Facebook," *Online Information Review*. Online Advanced Publication. <https://doi.org/10.1108/OIR-11-2021-0600>

- Brandtzaeg, P. B., Følstad, A., and Chaparro Domínguez, M. Á. (2018), “How journalists and social media users perceive online fact-checking and verification services,” *Journalism Practice*, Vol. 12 No. 9, pp.1109-1129. <https://doi.org/10.1080/17512786.2017.1363657>
- Brehm, J. W., and Brehm, S. S. (1981), *Psychological reactance: A theory of freedom and control*. San Diego, CA: Academic Press.
- Brehm, S. S., and Brehm, J. W. (2013), *Psychological reactance: A theory of freedom and control*. Academic Press.
- Byrne, S., and Hart, P. S. (2009), “The boomerang effect a synthesis of findings and a preliminary theoretical framework,” *Annals of the International Communication Association*, Vol. 33 No. 1, pp.3-37. <https://doi.org/10.1080/23808985.2009.11679083>
- Cacioppo, J. T., and Petty, R. E. (1984), The elaboration likelihood model of persuasion. *ACR North American Advances*.
http://130.18.86.27/faculty/warkentin/SecurityPapers/Merrill/CacioppoPetty1984_ELMPersuasion.pdf
- Census and Statistics Department. (2018, January), *Marriage and divorce trends in Hong Kong, 1991 to 2016*. <https://www.statistics.gov.hk/pub/B71801FB2018XXXXB0100.pdf>
- Chen, X. K., Na, J. C., Tan, L. K. W., Chong, M., and Choy, M. (2022), “Exploring how online responses change in response to debunking messages about COVID-19 on WhatsApp,” *Online Information Review*, Vol. 46 No. 6, pp.1184-1204. <https://doi.org/10.1108/OIR-08-2021-0422>
- Church, R., and Wilbanks, C. (1986), *Values and policies in controversy: An introduction to argumentation and debate*. (2nd ed.). Kendall Hunt Pub Co.
- Dillard, J. P., and Shen, L. (2005), “On the nature of reactance and its role in persuasive health

communication,” *Communication Monographs*, Vol. 72 No. 2, pp.144–168.

<https://doi.org/10.1080/03637750500111815>

Factcheck Lab. (2020), Fact-checking methods and standards. Factcheck Lab. Retrieved May 28, 2022, from <https://www.factchecklab.org/methodology/>

Frank, L. B., Murphy, S. T., Chatterjee, J. S., Moran, M. B., and Baezconde-Garbanati, L. (2015), “Telling stories, saving lives: creating narrative health messages,” *Health communication*, Vol. 30 No. 2, pp.154-163. <https://doi.org/10.1080/10410236.2014.974126>

Garrett, R. K., and Weeks, B. E. (2013, February), The promise and peril of real-time corrections to political misperceptions. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 1047-1058). <https://doi.org/10.1145/2441776.2441895>

Graves, L. (2018), “Boundaries not drawn: Mapping the institutional roots of the global fact-checking movement,” *Journalism Studies*, Vol. 19 No. 5, pp.613–631. <https://doi.org/10.1080/1461670X.2016.1196602>

Graves, L. Nyhan, B. and Reifler, J. (2016), “Understanding innovations in journalistic practice: A Field experiment examining motivations for fact-checking,” *Journal of Communication*, Vol. 66 No. 1, pp.102–138. <https://doi.org/10.1111/jcom.12198>

Graves, L., and Cherubini, F. (2016), *The Rise of Fact-Checking Sites in Europe* (Reuters Institute Digital News Report). Reuters Institute for the Study of Journalism. <https://www.mm.dk/wp-content/uploads/2016/11/The-Rise-of-Fact-Checking-Sites-in-Europe.pdf>

Gray, J. B., and Harrington, N. G. (2011), “Narrative and framing: A test of an integrated message strategy in the exercise context,” *Journal of Health Communication*, Vol. 16 No. 3, pp.264-281. <https://doi.org/10.1080/10810730.2010.529490>

- Greene, K., and Brinn, L. S. (2003), “Messages influencing college women's tanning bed use: Statistical versus narrative evidence format and a self-assessment to increase perceived susceptibility,” *Journal of health communication*, Vol. 8 No. 5, pp.443-461.
<https://doi.org/10.1080/713852118>
- Han, B., and Fink, E. L. (2012), “How do statistical and narrative evidence affect persuasion?: The role of evidentiary features,” *Argumentation and Advocacy*, Vol. 49 No. 1, pp.39-58.
<https://doi.org/10.1080/00028533.2012.11821779>
- Henke, J., Leissner, L., and Möhring, W. (2020), “How can journalists promote news credibility? Effects of evidences on trust and credibility,” *Journalism Practice*, Vol. 14 No. 3, pp.299-318.
- HKBU Fact Check. (2022), *Our verdicts and explanations*. HKBU Fact Check. Retrieved May 28, 2022, from <https://factcheck.hkbu.edu.hk/home/en/about-us/our-verdicts-and-explanations/>
- Hong, S., and Park, H. S. (2012), “Computer-mediated persuasion in online reviews: Statistical versus narrative evidence,” *Computers in human behavior*, Vol. 28 No. 3, pp.906-919.
<https://doi.org/10.1016/j.chb.2011.12.011>
- Immigration Department. (2015), *Annual report 2015*.
https://www.immd.gov.hk/publications/a_report_2015/en/ch1.html
- Kazoleas, D. C. (1993), “A comparison of the persuasive effectiveness of qualitative versus quantitative evidence: A test of explanatory hypotheses,” *Communication Quarterly*, Vol. 41 No. 1, pp.40–50. <https://doi.org/10.1080/01463379309369866>
- Kim, B., Cooks, E., and Kim, Y. (2021), “Thinking, checking and learning: Testing a moderated-mediation model of social media news use conditional upon elaboration on political knowledge via fact-checking,” *Online Information Review.*, Vol. 46 No. 5, pp.920-936.

<https://doi.org/10.1108/OIR-12-2020-0584>

Kim, S. Y., Levine, T., and Allen, M. (2013), “Comparing Separate Process and Intertwined Models for Reactance,” *Communication Studies*, Vol. 64 No. 3, pp.273–295.

<https://doi.org/10.1080/10510974.2012.755639>

Kreuter, M. W., Holmes, K., Alcaraz, K., Kalesan, B., Rath, S., Richert, M., ... Clark, E. M. (2010), “Comparing narrative and informational videos to increase mammography in low-income African American women,” *Patient Education and Counseling*, Vol. 81 No. Suppl, pp.S6-14. <https://doi.org/10.1016/j.pec.2010.09.008>

Lee, E., and Leets, L. (2002), “Persuasive storytelling by hate groups online: Examining its effects on adolescents,” *American behavioral scientist*, Vol. 45 No. 6, pp.927-957.

<https://doi.org/10.1177/0002764202045006003>

Mena, P. (2019), “Principles and boundaries of fact-checking: Journalists’ perceptions,” *Journalism Practice*, Vol. 13 No. 6, pp.657-672.

Murphy, S. T., Frank, L. B., Chatterjee, J. S., and Baezconde-Garbanati, L. (2013), “Narrative versus non-narrative: The role of identification, transportation, and emotion in reducing health disparities,” *Journal of Communication*, Vol. 63 No. 1, pp.116–137.

<https://doi.org/10.1111/jcom.12007>

Nabi, R. L., Moyer-Gusé, E., and Byrne, S. (2007), “All joking aside: A serious investigation into the persuasive effect of funny social issue messages,” *Communication Monographs*, Vol. 74 No. 1, pp.29-54. <https://doi.org/10.1080/03637750701196896>

Ng, K. (2018), If no ‘one-way’ permit scheme for migrants from mainland China, Hong Kong would have had 700,000 fewer workers over 20 years, labour chief Law Chi-kwong says.

Apple Daily. <https://www.scmp.com/news/hong-kong/society/article/2170567/without-one-way-permit-scheme-migrants-mainland-china-hong>

Nyhan, B., and Reifler, J. (2010), “When corrections fail: The persistence of political misperceptions,” *Political Behavior*, Vol. 32 No. 2, pp.303-330.

<https://doi.org/10.1007/s11109-010-9112-2>

Nyhan, B., and Reifler, J. (2011), Opening the political mind? The effects of self-affirmation and graphical information on factual misperceptions. Unpublished manuscript.

Ohlert, C. R., and Weißenberger, B. E. (2015), “Beating the base-rate fallacy: an experimental approach on the effectiveness of different information presentation formats,” *Journal of Management Control*, Vol. 26 No. 1, pp.51-80.

Pennycook, G., and Rand, D. G. (2019), “Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning,” *Cognition*, Vol. 188 No. September 2017, pp.39–50. <https://doi.org/10.1016/j.cognition.2018.06.011>

Pennycook, G., Bear, A., Collins, E. T., and 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*, Vol. 66 No. 11, pp.4944–4957. <https://doi.org/10.1287/mnsc.2019.3478>

Petty, R. E., and Cacioppo, J. T. (1986), The elaboration likelihood model of persuasion. In *Communication and persuasion* (pp. 1-24). Springer, New York, NY.

<http://fsnagle.org/files/petty1986elaboration.pdf>

PolitiFact. (2018), How we determine Truth-O-Meter ratings. *The Poynter Institute*.

<https://www.politifact.com/article/2018/feb/12/principles-truth-o-meter-politifact-methodology-i/#Truth-O-Meter%20ratings>

- Porter, E., Wood, T. J., and Kirby, D. (2018), “Sex trafficking, Russian infiltration, birth certificates, and pedophilia: A survey experiment correcting fake news,” *Journal of Experimental Political Science*, Vol. 5 No. 2, pp.159-164.
<https://doi.org/10.1017/XPS.2017.32>
- Rains, S. A. (2013), “The Nature of Psychological Reactance Revisited: A Meta-Analytic Review,” *Human Communication Research*, Vol. 39 No. 1, pp.47–73. <https://doi.org/10.1111/j.1468-2958.2012.01443.x>
- Shelby, A., and Ernst, K. (2013), “Story and science: how providers and parents can utilize storytelling to combat anti-vaccine misinformation,” *Human vaccines and immunotherapeutics*, Vol. 9 No. 8, pp.1795-1801.
- Shin, J., and Thorson, K. (2017), “Partisan Selective Sharing: The Biased Diffusion of Fact-Checking Messages on Social Media,” *Journal of Communication*, Vol. 67 No. 2, 233–255.
<https://doi.org/10.1111/jcom.12284>
- Snopes. (2021), Fact Check Ratings. *Snopes.com*. <https://www.snopes.com/fact-check-ratings/>
- Stencel, M. (2016), Global fact-checking up 50% in past year. *Duke Reporters’ Lab*, 16.
<https://reporterslab.org/global-fact-checking-up-50-percent/>
- Taiwan FactCheck Center. (2022), *Fact-checking verdict explanations*. HKBU Fact Check.
Retrieved May 28, 2022, from <https://tfc-taiwan.org.tw/node/5866>
- The Government of the Hong Kong Special Administrative Region. (2019), *Operation of One-way Permits scheme*. GovHK.
<https://www.info.gov.hk/gia/general/201903/20/P2019032000536.htm>

- Tsang, S. J. (2021), “Motivated fake news perception: The impact of news sources and policy support on audiences’ assessment of news fakeness,” *Journalism and Mass Communication Quarterly*, Vol. 16 No. 2022, pp.2260-2282. <https://doi.org/10.1177/1077699020952129>
- Tsang, S. J. (2022a), “Biased, not lazy: Assessing the effect of COVID-19 misinformation tactics on perceptions of inaccuracy and fakeness,” *Online Media and Global Communication*, Vol. 1 No. 3, pp.469-496.
- Tsang, S. J. (2022b), “Issue stance and perceived journalistic motives explain divergent audience perceptions of fake news,” *Journalism*, Vol. 23 No. 4, pp.823-840.
<https://doi.org/10.1177/1464884920926002>
- Tversky, A., and Kahneman, D. (1974), “Judgment under uncertainty: Heuristics and biases,” *Science*, Vol. 185 No. 4157, pp.1124-1131. <https://doi.org/10.1126/science.185.4157.1124>
- Vallone, R. P., Ross, L., and Lepper, M. R. (1985), “The hostile media phenomenon: biased perception and perceptions of media bias in coverage of the Beirut massacre,” *Journal of Personality and Social Psychology*, Vol. 49 No. 3, pp.577-585.
<https://doi.org/10.1037/0022-3514.49.3.577>
- Walter, N., and Murphy, S. T. (2018), “How to unring the bell: A meta-analytic approach to correction of misinformation,” *Communication Monographs*, Vol. 85 No. 3, pp.423–441.
<https://doi.org/10.1080/03637751.2018.1467564>
- Walter, N., Cohen, J., Holbert, R. L., and Morag, Y. (2020), “Fact-checking: A meta-analysis of what works and for whom,” *Political Communication*, Vol. 37 No. 3, pp.350-375.
<https://doi.org/10.1080/10584609.2019.1668894>
- Walter, N., and 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*, Vol. 47 No. 2, pp.155-177.

Yoel, R., and Nick, P. (2020, May 11), *Updating our approach to misleading information*.

https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html

Young, D. G., Jamieson, K. H., Poulsen, S., and Goldring, A. (2018), “Fact-checking effectiveness as a function of format and tone: Evaluating FactCheck.org and FlackCheck.org,”

Journalism and Mass Communication Quarterly, Vol. 95 No. 1, pp.49-75.

<https://doi.org/10.1177/1077699017710453>

Zillmann, D., and Brosius, H. B. (2000), *Exemplification in communication: The influence of case reports on the perception of issues*. Mahwah, NJ: Lawrence Erlbaum.

Figure 1.

Descriptive statistics for message discounting (perceived news fakeness) by evidence type and veracity

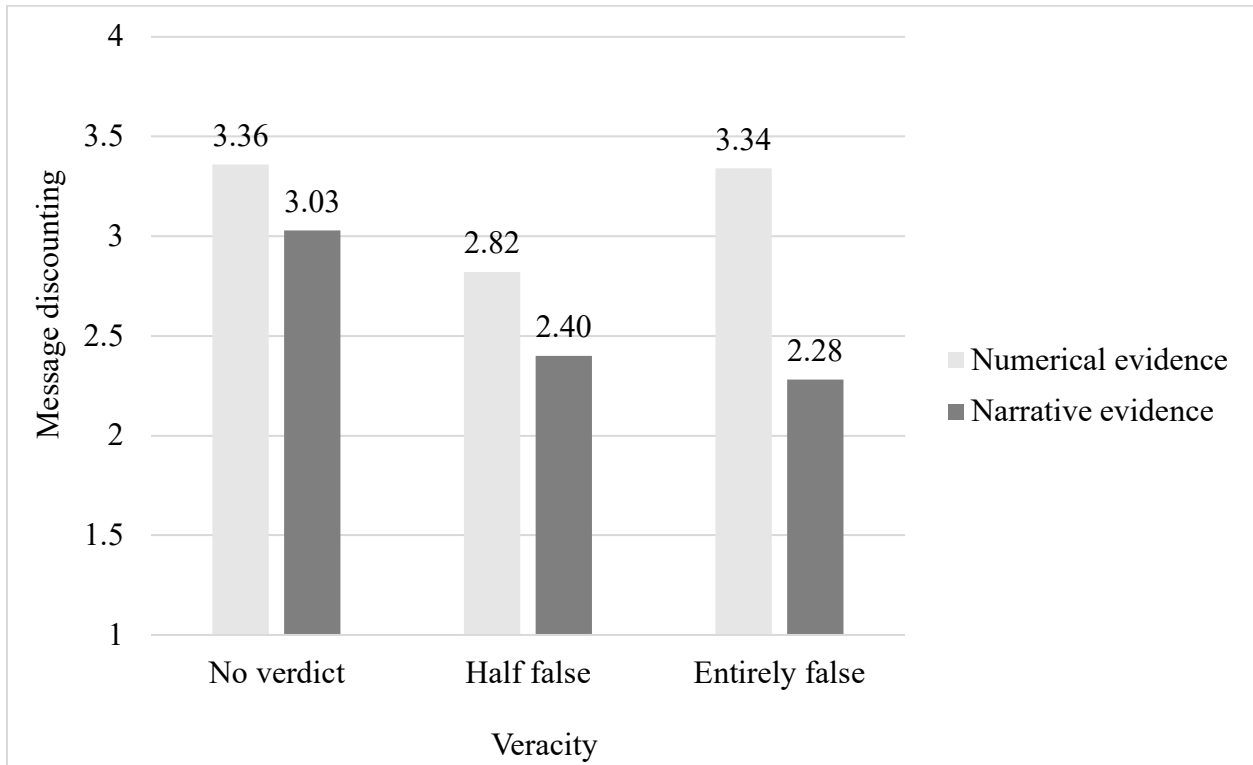


Figure by authors

Figure 2.

Descriptive statistics for policy support (policy evaluation) by evidence type and veracity among participants who disagreed the news article

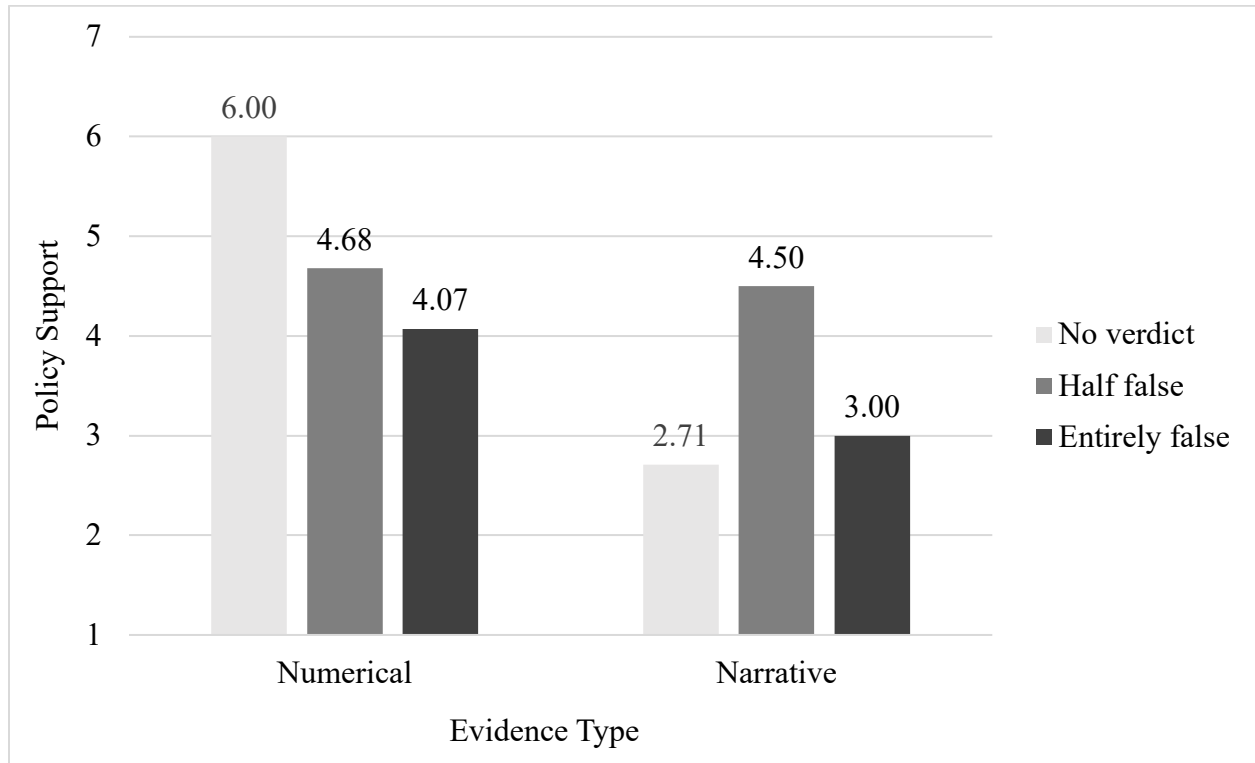


Figure by authors