Title
How Do Language Intensity and Artificial Intelligence (AI) Affect Perceptions of Fact-checking Messages and Evaluations of Fact-checking Agencies?

Author
Xue, Haoning

Publication Date
2021

Peer reviewed|Thesis/dissertation

How Do Language Intensity and Artificial Intelligence (AI) Affect Perceptions of Fact-checking Messages and Evaluations of Fact-checking Agencies?

By

HAONING XUE

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF ARTS

in

COMMUNICATION

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

____
Jingwen Zhang, Chair

____
Cuihua Shen

____
Magdalena Wojcieszak

Committee in Charge

2021

Abstract

Fact-checking agencies are essential to correct misinformation and inform the public, while how people evaluate these agencies and their messages remain unclear (Brandtzaeg et al., 2017). Two factors about the messages and the sources – two essential factors in the theories of persuasion – were examined: language intensity of fact-checking labels and AI as a fact-checking agency. Language intensity, a linguistic feature that reflects message specificity and emotionality, may implicitly influence the acceptance of misinformation corrections and behavior intentions (Bowers, 1963). While AI has the potential to automate the fact-checking process and improve the acceptance of misinformation corrections as an unbiased automated decision maker, the social acceptance of AI in fact checking is unclear. This study investigated how language intensity and fact-checking agency (human vs. AI) influence the evaluations of fact-checking messages and agencies with an observational study of fact-checking messages on social media (N = 33755) and two online experiments (combined N = 1449) in the U.S. Both the observational study and the experiments showed that fact-checking messages with high language intensity would elicit low message credibility, while this effect diminished when the messages were counter-attitudinal in the experiments. Besides, participants perceived AI fact-checking agencies the same as human agencies. Individual differences in conspiracy ideation, political ideology and demographics significantly affected message credibility and engagement intentions as well. These findings suggest that language nuances such as language intensity in fact-checking messages affected message perception and the acceptance of misinformation corrections. Theoretical and practical implications were discussed in detail.

Introduction

Misinformation creates and intensifies public misperceptions and behaviors in the realm of politics, health, and other social issues (Walter & Murphy, 2018). Especially technological advances enable the faster transmission of misinformation on social media without traditional media gatekeepers (Shin, 2021). The effect of misinformation can linger and even strengthen after exposure to corrections (Thorson, 2016). However, fact-checking, a systematic validation of public claims, is promising to effectively alleviate the damaging effect of misinformation circulated on social media (Walter et al., 2019). Fact-checking is not a novel form of journalism (Grant, 2000), but independent fact-checking agencies (e.g., snopes.com) and in-house fact-checking groups (e.g., AP Fact Check) only emerged in recent years to cope with the viral spread of misinformation.

Fact-checking agencies commonly use fact-checking labels such as true, false, and mostly false to convey their decisions on checked misinformation in a straightforward manner (e.g., Truth-O-Meter by PolitiFact). Since people mostly engage in heuristic processing of information on social media (Chaiken & Ledgerwood, 2012), fact-checking labels – commonly emphasized with various visual cues – are crucial heuristics of persuasion (Oeldorf-Hirsch et al., 2020). The persuasiveness of fact-checking labels are usually accomplished by conveying message strength through language intensity, a linguistic feature that reflects message specificity and emotionality (Bowers, 1963; Hamilton et al., 1990). This is a subtle aspect that may implicitly influence attitude and is positively associated with persuasiveness (Burgers & de Graaf, 2013). Although fact-checking agencies are expected to be and usually perceived as unbiased as a part of journalism, there is skepticism about the objectivity and transparency of those agencies from both professional journalists and social media users (Brandtzaeg et al., 2017). Expressions with different levels of language intensity, such as different levels of emotional arousal conveyed by *fantastic* and *nice*, may imply partisanship in journalism practices by inducing a reporting bias, which suggests a perception that facts are not accurately reported (Burgers & de Graaf, 2013; Eagly et al., 1978). In fact-checking contexts, it is possible that different levels of language intensity exist in fact-checking labels by describing misinformation untruthfulness. For instance, the implied untruthfulness varies when two equally false statements are labeled as *wrong* or *misleading*. The processing and effects of such subtle linguistic variations in fact-checking remain unexamined.

In addition to the fact-checking labels, fact-checking agencies matter significantly as the message sources and influence the perceived trustworthiness (e.g. Zhang et al., 2021). AI has been increasingly assisting with the fact-checking process and has the potential to improve the acceptance of misinformation corrections with an unbiased and non-partisan image (Sundar, 2008), while little is known about the social acceptance of AI as fact-checking agencies. Social media platforms (e.g., Facebook), technology companies (e.g., fullfact.org), and academia (e.g., Hassan et al., 2017; Karadzhov et al., 2017) have endeavored to accelerate the fact-checking process through automated fact-checking conducted by AI. Currently, AI mostly assists humans in the decision-making process as full automation is not yet achieved. Still, people tend to perceive AI as more objective than human (Sundar, 2008), which implies that people might be more willing to keep an open mind to counter-attitudinal corrections. Even though AI is perceived with great potential, algorithm aversion exists, and the public perception of AI is context specific. For instance, AI is criticized for being biased in gender and race (e.g., Dastin, 2018) and flawed in areas such as automated driving (e.g., Conger, 2020). Since how people perceive AI fact-checking is still overlooked, the mixed evidence makes it necessary to investigate whether the perception of misinformation corrections differs when fact-checking is done by humans or AI.

The impact of language intensity of fact-checking labels and AI on belief updating and behavior change remains understudied in the fact-checking contexts. Limited work investigates the perception of fact-checking agencies or language intensity in this context, especially in terms of perceived credibility, bias, and trustworthiness. The proposed study fills the gap by addressing two fundamental questions: 1) Does language intensity of fact-checking labels affect people's trust in fact-checking messages and fact-checking agencies? 2) Do people perceive fact-checking agencies and messages differently when the fact-checking is conducted by humans or AI? The theoretical argument is grounded on the theoretical frameworks of an information processing model of language intensity (Hamilton, 1998) and machine heuristic (Sundar, 2008).

This study relies on a combination of an observational study of fact-checking posts on social media (N = 33755) and two factorial experiments (combined N = 1449) to investigate the effect of language intensity and fact-checking agency across three different topics in politics, health, and economics. This project contributes to the literature of language intensity and message effect by extending to the fact-checking contexts in both real and laboratory settings. It provides a new perspective to understand the conditions under which people are more willing to accept counter-attitudinal misinformation corrections and to understand social acceptance of AI. This study also provides practical implications in language preciseness and AI application for journalism practices in the mediated contexts. The theoretical frameworks used in this study together with empirical evidence are revisited below; hypotheses, research questions, and results of all three studies are reported afterwards.

Language Intensity and persuasion

Language intensity is a linguistic feature that reflects message specificity and emotionality, which is associated with language extremity in evaluations (Bowers, 1963, as cited in Burgers & de Graaf, 2013; Hamilton et al., 1990). Specificity refers to the degree of language concreteness, e.g., *sexual assault* is more detailed than *violence*, while emotionality suggests that language varies across the degree of emotion in evaluative statements, e.g., *fantastic* is perceived as more positive than *nice*, and *ruthless* is perceived as more negative than *unkind* (Burgers & de Graaf, 2013).The focus of the fact-checking contexts is language emotionality and extremity; language variations in fact-checking exist across the spectrum of message untruthfulness through lexical and semantic intensifiers (Athanasiadou, 2007). For example, *misleading* and *deceptive* can be used to describe the same misinformation while implying different levels of untruthfulness intentions: *deceptive* is usually perceived as intentionally false and therefore more negative than *misleading*. Besides, *mostly false* is perceived as more true than *completely false*. These adjectives are commonly used by fact-checking agencies to label or describe misinformation. How fact-checking agencies attach different labels with different issues or entities may imply different levels of language intensity and create a clear but subtle partisanship (Abelson, 1995).

Language intensity is associated with evaluations of journalistic neutrality and news quality as a form of news sensationalism (Burgers & de Graaf, 2013). Though sensationalism in news sometimes is appreciated for a human touch, news media are usually expected to be neutral. This expectation also applies to fact-checking as a form of journalism, and fact-checking agencies are often perceived as factual and unbiased (Brandtzaeg et al., 2017). Such attitudinal extremity conveyed by language intensity in fact-checking labels can be perceived as a violation of journalistic neutrality, which is negatively associated with evaluations of newsworthiness and trustworthiness. This tendency is theorized as a reporting bias, “the belief that a communicator's willingness to convey an accurate version of external reality is compromised” (Eagly et al., 1978). Therefore, fact-checking labels with high language intensity that deviates from neutrality might not only harm message evaluation but damage source credibility as well. This connection is also supported by the information processing model of language intensity that establishes the effect of language intensity on source and message evaluation (Hamilton, 1998).

Although little empirical evidence exists regarding the effect of language intensity in the fact-checking contexts, empirical evidence confirmed a negative effect on message and source evaluation to some extent. Burgers and de Graaf (2013) found a negative effect of language intensity on newsworthiness, though it became a positive indirect effect with a different topic in the second study. When a mild transgression was exaggerated as a scandal in political news, participants held lower levels of message appropriateness and trust; when it came to a severe transgression, scandalization made no difference (Graβl et al., 2019). In other cases, however, language intensity has been found to increase persuasiveness (e.g., Clementson et al., 2015; Hamilton & Stewart, 1993) or have no effect (Bowers, 1963).

There are two possible explanations for this discrepancy. First, the effect of language intensity is subject to a variety of contextual factors such as operationalization of language intensity (Liebrecht et al., 2019) and topic relevance (e.g., Burgers & de Graaf, 2013). Changes in the contexts bring in external factors that might alter the main effect. High language intensity might be encouraged in certain occasions (e.g., political speech) while discouraged in others. Topic relevance matters as well; language intensity may not have any effect when people do not care about a matter. Second, other variables involved might not be identified, such as source credibility, argument quality, and gender (Hamilton, 1998). These factors may act as mediators or moderators in the process. Therefore, the influence of language intensity remains unclear in the fact-checking context.

Artificial Intelligence in the Fact-checking Contexts

News organizations and fact-checking agencies rely on professional journalists to produce factual and non-partisan information. However, misinformation disseminated on social media has been soaring over the last decade, which makes it difficult for professionals to monitor, detect, and correct misinformation on social media. Therefore, AI starts to play an assisting role in misinformation detection and fact-checking. It detects information truthfulness based on existing ground truth, linguistic features, and social media engagements (e.g., Hassan et al., 2017). Since AI cannot fully capture the linguistic and factual nuance that may determine information truthfulness, fully automated fact-checking AI is under development by researchers and technology companies (e.g., Fullfact.org, Logically.ai). Currently, major technology companies also automatically take down misinformation on social media that have been identified by fact-checking agencies to stop further spread (Facebook, 2020). AI shows the potential to fully automate the laborious fact-checking process and may play a major and independent role in future journalism practice, but the social acceptance of AI is unclear and the research on the perceptions of AI fact-checking is necessary.

There are substantial definitions and categories of AI in the literature; the definition of AI also evolves over time and technology development (Wang, 2019). A recent definition by (Ginsberg, 2012) states that AI is “the enterprise of constructing an artifact that can reliably pass the Turing test”, which implies that AI acts like a human and is as intelligent as or even more intelligent than a human. Terms such as algorithm, machine, and automated decision maker point at the notion of human-like capability and intelligence. In the context of this study, AI refers to a non-human enterprise or machine that performs the same task as humans do. AI fact-checking agents would engage in tasks such as identifying information check-worthiness and truthfulness and posting fact-checking messages on social media.

Literature suggests mixed theoretical frameworks and empirical findings regarding the perception of AI versus humans. In the context of interpersonal communication, (Nass and colleagues (1994) proposed that individuals perceive computers as independent social entities and apply the rules of human interactions to computers in the theoretical framework of Computers are Social Actors (CASA). This claim suggests that humans treat humans and AI without difference, which is based on the assumption that humans expect AI to be a similar social actor like humans. However, this assumption may not hold currently as individuals have increasingly more direct interactions with AI over the last two decades, which leads to updated perceptions and expectations of AI (Gambino et al., 2020).

This discrepancy in expectation is captured by the Modality-Agency-Interactivity-Navigability (MAIN) model, in which Sundar (2008) referred to the tendency that individuals perceive machines as more objective and non-partisan than humans as *machine heuristic*. This mental shortcut implies a positive stereotype about “machine infallibility and neutrality” (Sundar & Kim, 2019) that is gradually formed with more interactions with and knowledge on machines, which is related to a mindless trust in automation and automation bias under the theoretical framework of heuristic processing (Mosier et al., 1998). When applied to the fact-checking contexts, machine heuristic indicates that individuals are more likely to develop more favorable perceptions and behavior intentions when it comes to AI fact-checkers.

Algorithm appreciation describes this mindless preference for AI even when people have limited knowledge on the underlying mechanism of AI (Logg et al., 2018, as cited in Wojcieszak et al., 2021). In addition to the study context of numeric estimation task, empirical findings have supported the preference for AI over humans and even experts in privacy revealing (Sundar & Kim, 2019), recruiting process (Hong et al., 2020), and general topics in health, media, and justice (Araujo et al., 2020). However, a recent meta-analysis of 11 experiments published in the last three years found that there is no significant difference in perceptions of automated and human-written news, while human-written news is perceived as high-quality and well-written in general (Graefe & Bohlken, 2020). This finding might be explained by the fact that current AI is not fully intelligent in imitating natural expressions and news writing; therefore, the favorability toward humans might attribute to the preference for quality news.

But one boundary condition of algorithm appreciation identified is algorithm failing: algorithm aversion occurs when people see AI make mistakes and perform worse than humans, since people tend to assume AI to be perfect as the machine heuristic indicates (Dietvorst et al., 2014). Since the contexts of these two studies by Logg and colleagues (2018) and Dietvorst and colleagues (2014) involved numeric estimation tasks, the assumption of machine infallibility might be reasonable as machines have greater advantages in big data access and processing than humans. Automated driving is another context in which machine failures were more unacceptable than human's (Hong, 2020). But when it comes to humanity issues such as automated identification of hate speech, people are more forgiving of AI than human (Shank et al., 2019). People also trust human more when it came to online content moderation (Wojcieszak et al., 2021). This is also consistent with the idea of expectation fulfillment. Therefore, when AI fails to meet this expectation, people would show averse attitudes toward AI. As machine heuristic predicts machine objectivity, we argue that people would expect AI to be unbiased and the preference for AI would exist in the fact-checking context, although the effect of fact-checking agency on behavior intentions is unclear.

Study 1: Observational Study

Since social media have become a major source of news, social media platforms provide a platform to observe the natural occurrences of and connections between different constructs. This observation study of posts and comments from seven US-based fact-checking agencies on Facebook (N = 33755; March 2010 – January 2021) was conducted to investigate (a) whether variations of language intensity exist in fact-checking labels and (b) the potential association between the language intensity of fact-checking labels and the engagement with the posts (i.e., likes, comments, and shares).

Both computational methods (e.g., natural language processing) and pilot study with human participants were used to quantify language intensity and emotionality (as an aspect of intensity). Since the attitude toward fact-checking issues can be hard to capture on social media, the main dependent variable is the actual engagement. Since there is currently no AI fact-checking agency operating, fact-checking agency was not included as a factor in this case. This section proposes hypotheses related to the research question, followed by a detailed description of data collection and processing, results reporting, and discussion.

Language intensity, emotionality and engagement

Language intensity can be positively associated with behavior change since it indicates larger message strength (Craig & Blankenship, 2011). Previous findings showed that emails using intense language elicited a higher response rate (Andersen & Blackburn, 2009); people were more likely to engage in health behaviors with more intense language (Buller et al., 2000, 2009). It is possible that language intensity increases message engagement as well, since message engagement on social media can be affectively and cognitively triggered (Kim & Yang, 2017). However, there is little evidence on its effect on message engagement, such as likes, comments, and shares. Therefore,

RQ1. What is the association between language intensity of fact-checking messages and the number of likes, comments, and shares of fact-checking messages?

Since emotionality is one of the aspects of language intensity, it is expected that emotionality is correlated with language intensity and possesses a positive effect on engagement as well.

H1. Emotionality of fact-checking messages would be positively associated with the number of likes, comments, and shares of fact-checking messages.

Data and Methods

Data collection. First, Facebook was chosen because fact-checking agencies tend to have more engagements and attention on Facebook than other social media platforms such as Twitter. Further, fact-checking related keywords on Facebook with a brand-new Facebook account were searched to retrieve Facebook pages of all relevant fact-checking agencies. Relevant agencies refer to those that (a) operate in the U.S., (b) report in English, (c) report original fact-checking stories, (d) falls into the category of *News & Media Website*, and (e) has a verified badge. This search results in seven fact-checking agencies. Second, Facebook's CrowdTangle data monitoring platform was used to collect all historical posts (N = 88,598; CrowdTangle Team, 2021). Not all posts were fact-checking posts; irrelevant ones included website donation and repost of different websites. Since almost all posts were attached with a link to the official website of a fact-checking agency, irrelevant posts were removed by identifying features in the website links. For example, AFP fact-checking articles were featured by "*factcheck.afp.com*"; fact-checking articles by Snopes.com were archived by years in the links, from 2016 to 2021. Therefore, irrelevant posts were removed automatically with Python, resulting in 33,755 fact-checking posts. Among these posts, 29,374 posts have received at least one comment, and all comments attached were collected with Facepager (N = 1,284,813; Jünger & Keyling, 2019). It is worth noticing that the number of comments retrieved does not equal the number of comments made, since some comments may have been deleted or have privacy settings.

Linguistic features of language intensity. To my best knowledge, there is no corpus on language intensity in untruthfulness. However, one aspect of language intensity is emotionality (Burgers & de Graaf, 2013), which can be captured with existing natural language processing tools such as IBM Watson Natural Language Understanding (IBM Watson, 2021). With IBM Watson, we were able to obtain the sentiment score for each post ranging from -1 (negative emotion) to 1 (positive emotion). Since emotionality does not consider emotional valence, emotionality was represented by the absolute value of the sentiment score. All comments attached to a post were aggregated to a single text for convenient processing. Comments of 38 posts were unable to be processed by IBM Watson due to errors in length or special characters, resulting in 29,336 sets of comments and 33,755 posts with emotionality scores.

Pilot study of language intensity. To my best knowledge, there is no corpus on language intensity in untruthfulness. Therefore, 30 most frequently used adjectives describing information untruthfulness from these fact-checking posts were collected (e.g., *wrong*, *misleading*, *unfounded*, *manipulated*). Addiction applies when a post with multiple veracity-related adjectives or with one adjective appearing more than once, since more adjectives used indicate higher levels of language intensity. 24% of all posts contain at least one of these adjectives (*N* = 7,962).

Language intensity of these 30 adjectives were determined by a sample of undergraduate students enrolled in a communication introduction course from a public university in California. To make sure that participants have similar levels of English proficiency and experience, 43 responses completed by international students were excluded, and the final sample size is 191. In the pilot study, participants were asked to reflect on 10 random adjectives out of 30 based on their daily experiences and to rate how intense these adjectives are.

Measures

Perceived intensity. The intensity of an adjective was measured by 10 items from the Language Intensity Scale on a 7-point scale (Hamilton & Stewart, 1993). This scale included three universal semantic features – evaluation (e.g., emotional – unemotional), potency (e.g., potent – impotent), and activation (e.g., active – inactive; Osgood, 1969). These 30 adjectives were generally intense (*M* = 4.27; *SD* = 1.34), with 21 of 30 being rated above four (Table 1).

Perceived intensity of each fact-checking post was measured by the number of times that each adjective appeared in this post multiplied by the corresponding intensity score of that adjective. Among all posts, 24% of them contained more than one of the veracity-related adjectives pretested (N = 7962; *M* = 4.92, *SD* = 1.76). Taking all posts into consideration, the fact-checking posts were not highly intense (N = 33755; *M* = 1.16; *SD* = 2.26; Table 2).

Engagement. The engagement of a fact-checking post (*M* = 748.22; *SD* = 1980.40; Table 2) was measured by the sum of the number of likes, comments, and shares of a post. This statistic was obtained through CrowdTangle (2021).

Emotionality. Emotionality, an important and measurable aspect of language intensity, was represented by the absolute value of the sentiment score since emotional valence does not matter in the case of emotionality. Both posts (*M* = 0.54, *SD* = 0.33) and comments (*M* = 0.57, *SD* = 0.18) were moderately emotional (Table 2).

Results

Descriptive statistics of fact-checking agencies. Though the oldest fact-checking post in this dataset dated back to 2010, there was a significant soar of fact-checking messages since 2018. Fact-checking posts created in the last four years (N = 27614) contributed to 82% of this dataset. Among 7 fact-checking agencies included, most agencies were active with more than 2,000 fact-checking posts, except *Health Feedback* (N = 32). *FactCheck.org* was the most popular agency with more than one million comments in total and approx. 3,700 engagement per post (*M* = 3777.18; *SD* = 4370.40). The fact-checking posts from *AFP Fact Check* (*M* = 3.93, *SD* = 2.37; $t = 73.15$, $p < .001$), *FactCheck.org* (*M* = 2.82, *SD* = 3.23; $t = 30.60$, $p < .001$), and *Health Feedback* (*M* = 2.99, *SD* = 4.03; $t = 3.18$, $p < .01$) were significantly more intense than those from other four agencies (*M* = 0.72; *SD* = 1.83). An average fact-checking post induced approximately 412 likes, 122 comments, and 214 shares.

Descriptive statistics of language intensity. On average, the language intensity of a fact-checking post was relatively low on a 7-point scale (*Min* = 0, *Max* = 19.22; *M* = 1.16, *SD* = 2.26). The most frequent adjectives were *false* (*M* = 4.19; *SD* = 1.34; *N* = 4047), *misleading* (*M* =

4.32; *SD* = 1.16; *N* = 853), *fake* (*M* = 4.65; *SD* = 1.33; *N* = 762), and *wrong* (*M* = 4.48; *SD* = 1.38; *N* = 604). The intensity of these words tended to be moderate. The most intensive adjectives tended to describe intentionally false information, such as *manipulated* (*M* = 5.60; *SD* = 1.18), *fraudulent* (*M* = 5.35; *SD* = 1.04), *deceptive* (*M* = 5.12; *SD* = 1.04), and *inflammatory* (*M* = 5.12; *SD* = 1.18). U*nclear* was considered the least intense (*M* = 2.93; *SD* = 1.19; *N* = 88). Language intensity of fact-checking posts was positively correlated with engagement (r = .11, p < .001) and emotionality in posts (r = .17, p < .001) and comments (r = .06, p < .001). It suggested that higher levels of language intensity in posts were positively associated with higher levels of engagement and emotionality in comments, and that emotionality represented some aspect of language intensity.

To test H1 and H2, Poisson regression models were conducted since intensity scores and engagement had Poisson distributions (Table 3). The number of followers of fact-checking agencies, post word counts, and post emotionality were controlled in both models. Robust standard errors were used due to a large sample size and homoscedasticity to obtain robust p-values. The results showed that language intensity was a positive indicator of engagement with fact-checking posts (b = 0.08, p < .001) and emotionality in comments (b = 0.01, p < .001). Specifically, the effect of language intensity was consistent for the number of likes (b = 0.03, p < .001), comments (b = 0.04, p < .01), and shares (b = 0.06, p < .001). It suggested that social media users were more likely to engage with fact-checking posts and to become more emotional in the comments with higher levels of language intensity in the fact-checking posts. Therefore, H1 was supported.

Not surprisingly, emotionality in posts and comments was positively associated with each other (b = 0.04, p < .001): the more emotional a fact-checking post is, the more emotional its comments are. However, emotionality in posts negatively predicted engagement (b = -0.12, p = .003), with more emotional posts attracting less engagement from the audience. However, a significant negative association was only found for the number of comments (b = -0.15, p < .001), while nonsignificant effects were found for the number likes (b = -0.06, p = .121) and shares (b = -0.05, p = .482). H2 was not supported.

Discussion

This observational study of fact-checking posts and comments on social media provides initial evidence that language intensity matters in the fact-checking contexts. Overall, fact-checking posts possessed a relatively low level of language intensity and fact-checking agencies maintain a high level of objectivity. Specifically, when it came to fact-checking posts that used words with higher levels of language intensity, people not only engaged with more likes, comments, and shares, but were more emotional in their comments as well. It may suggest that language intensity serves as both affective and cognitive triggers of social media engagement (Kim & Yang, 2017). This finding extends the study of language intensity by building a concrete connection between language intensity and message engagement.

Though emotionality is one significant aspect of language intensity, emotionality had a distinct effect on engagement. Overall, a negative effect on engagement was found, but a significant negative effect was only found for the number of comments, not likes and shares. It suggested that emotionality of fact-checking posts suppressed expressions of opinions, though it had no effect on the liking and sharing behaviors. This might be explained by the fact that emotionality accounts for a portion of language intensity, and the extremity part of language intensity might mainly contribute to its positive effect on engagement and the “attitude-behavior correspondence” (Craig & Blankenship, 2011).

However, this observational study has a few limitations. First, this study may not represent the population since it only covered the audience of fact-checking agencies on social media and they are likely to be their followers, while fact-checking agencies are not a mainstream media that the majority of the population are exposed to. In the future study, this could be addressed with a larger scale of fact-checking messages from any news media, not limited to independent fact-checking agencies. Further, similar to most observational studies, it did not establish a causal inference of the effect of language intensity. Lastly, the influence of individual characteristics remains unexplored since it is usually a violation of privacy to acquire personal information on social media. To address these issues, an online experiment was conducted to further explore the effect of language intensity, fact-checking agency, and individual characteristics.

Experiment 1

The observational study has confirmed that variations of language intensity exist in fact-checking messages, and that language intensity positively indicated message engagement. This experiment aims to examine the influence of language intensity and fact-checking agencies (AI vs. human) on source and message evaluations and establish a causal inference in a laboratory environment. This section presents the hypotheses and research questions in this study, experiment procedure and design, as well as a detailed report on results and discussion.

Language intensity and persuasion. Empirical evidence suggested that the persuasive effect of language intensity is context-specific, though the reporting bias predicts that high language intensity decreases source credibility (Eagly et al., 1978). Therefore,

RQ1. How does language intensity of fact-checking messages affect perceived credibility of fact-checking messages and engagement intentions?

AI in the fact-checking context. Since fact-checking is a part of journalism where people expect accuracy and objectivity (Brandtzaeg et al., 2017), machine heuristic that emphasizes on machine neutrality and nonpartisanship fits the context better (Sundar, 2008). Machine heuristic indicates a favorable attitude toward AI over human, though there is no prediction on engagement intentions. Therefore,

H1. Individuals would have better perceived credibility on fact-checking messages toward AI than human fact-checking agencies.

H2. Individuals would have better evaluations of fact-checking agencies toward AI than human fact-checking agencies.

RQ2. Do engagement intentions vary across fact-checking agencies?

Effects of Motivated Reasoning

Individual motives lead to selective information processing and biased evaluation (Chaiken & Ledgerwood, 2012; Kunda, 1990). Driven by confirmation and disconfirmation biases, people tend to favor information that is consistent with existing beliefs and worldviews and criticize counter-attitudinal information (Taber & Lodge, 2006). Empirical evidence suggests that motivated reasoning exists in the contexts of politics (Thorson, 2016), health (Bode & Vraga, 2017), climate change (Hart & Nisbet, 2012), and emerging technologies (Druckman &

Bolsen, 2011). People are also motivated to process misinformation corrections, in which people are more reluctant to accept counter-attitudinal corrections as true and persuasive (Walter et al., 2019). Motivated reasoning can be influenced by linguistic features. Language intensity was found to be negatively associated with attitude toward the source when it comes to counter-attitudinal information, while language intensity has no effect on attitude when the information is pro-attitudinal (Hosman, 2002). Still, the research on language intensity and motivated reasoning is limited, and it is difficult to make predictions at this point. Therefore,

RQ3. When fact-checking messages are counter-attitudinal, how does language intensity affect perceived credibility on fact-checking messages and engagement intentions?

As for the prior effect on fact-checking agency, the literature is also thin. Machine heuristic does imply that machines tend to be associated with higher objectivity and therefore, credibility (Sundar, 2008). A relevant study (Zarouali et al., 2020) found that people tended to agree more with counter-attitudinal news when it is delivered by chatbots than online news websites, and one of the influencing factors is credibility. In the context of online moderation, AI did not make people more open-minded (Wojcieszak et al., 2021). Therefore,

RQ4. When fact-checking messages are counter-attitudinal, how does fact-checking agency affect perceived credibility on fact-checking messages and engagement intentions?

Methods

Study design. To investigate the influence of language intensity and fact-checking agency, a factorial experiment with a mixed design was conducted. The between-subject factor in this experiment was the fact-checking agency, where participants read messages from either human or AI fact-checking agency. Since people tend to encounter messages of varied language intensity instead of the same level of intensity on social media, language intensity (low, mid, high) was included as a within-subject factor to ensure a high level of external validity. High, mid, and low levels of language intensity were represented by three adjectives respectively – *fake*, *wrong*, and *inaccurate*. These three adjectives were selected because (a) the meanings are all related to untruthfulness only, without implying the intention of untruthfulness, (b) they were used relatively commonly in the fact-checking messages, and (c) they varied across language intensity.

Another within-subject factor was the fact-checking topics, including the presidential election fraud in 2020, Covid-19 vaccine, and the raise of federal minimum wage. These topics were chosen as they had a prominent salience and representativeness in recent fact-checking messages. Each participant randomly read three fact-checking messages that cover all possible combinations of three levels of language intensity (one random adjective for each level) and three topics, with each message mapped with a random language intensity and a random topic. Political affiliation and ideology, AI familiarity, topic-specific prior attitude, conspiracy ideation, and demographics were included as covariates.

Sample and procedure. American participants ($N = 657$) were recruited from the SONA system of a public university in California in exchange for 0.5 research credit. Participants were redirected to a Qualtrics survey once they accepted the task ($N = 824$). After signing the consent, participants answered a few questions on their familiarity with AI and their attitudes on three topics. Further, participants were randomly assigned to read three fact-checking messages by an AI or human fact-checking agency. For each message, participants were asked to answer questions about their perceptions on this message. They were able to review the message while answering these questions. After reading all three messages, participants were asked to evaluate the manipulated agency. This online experiment ended with questions on conspiracy ideation and demographics to avoid the potential priming effect.

Non-American responses (N = 155) and blank responses were removed (N = 12). Therefore, the final sample was reduced to N = 657 (132 male, 504 female). Participants were aged from 18 to 67 (M = 20.32, SD = 2.85) and well educated with most having some college experience (N = 461). Participants' median annual household income fell in the \$80,000 to \$99,999 range. A detailed summary of sample characteristics is summarized in Table 4. There was no significant difference in age, gender, or education across conditions, though the income level was significantly higher in the AI condition than in the human condition (M_{AI} = 6.96, M_{human} = 6.18, p = .035). Since the income was not associated with the major findings, no adjustment was made.

Stimuli. In total, three fact-checking Facebook messages were created based on recent fact-checking messages on social media, ranging from October 2019 to January 2021. The format of a Facebook post was chosen because fact-checking agencies tend to have more engagements and attention on Facebook than other social media platforms such as Twitter. Each message consisted of 1) the fact-checking section: the claim that [a misinformation statement] was [manipulated adjective], such as *the claim that the covid vaccine is harmful was false*, 2) the explanation of the fact-checking decision, and 3) the credit section that states the fact-checking agency and post writer to be professional journalist or AI (see Figure 5). Each message contained one neutral topic-relevant image covered by the manipulated adjective, which was congruent with the practices of most fact-checking posts on social media (i.e., snopes.com). Consistent with the experiment conditions, the username was either *FactChecker* or *FactCheckingAI*. Other aspects were kept constant, such that the social media metrics were the same as the median of fact-checking posts on Facebook.

Measures

Message intensity. On a 7-point semantic differential scale, message intensity (*M* = 4.74, *SD* = 1.07, McDonald's ω = .84) was measured by seven items from an existing language intensity scale, including weak – strong, hesitant – emphatic, uncertain – certain, non-opinionated – opinionated, mild – intense, not extreme – extreme, powerless – powerful (Burgers & de Graaf, 2013; Hamilton & Stewart, 1993).

Message credibility. The message credibility (*M* = 5.04, *SD* = 1.21, McDonald's ω = .88) was measured by six items on a 7-point semantic differential scale, including not sensationalistic – sensationalistic, biased – unbiased, unfair – fair, non-factual – factual, non-objective – objective, and inaccurate – accurate (Sundar, 2008).

Evaluation of fact-checking agencies. Evaluation of fact-checking agencies (*M* = 4.91, *SD* = 1.24, McDonald's ω = .96) was concerned with three dimensions from an existing measure of evaluation of computer agent on a 7-point semantic differential scale (Brave et al., 2005): likability (four items, e.g., unpleasant – pleasant), trustworthiness (four items, e.g., dishonest – honest), and intelligence (three items, e.g., incapable – capable).

Engagement intention. On a 7-point scale (1 – disagree, 7 – agree), the intention to engage with the fact-checking message (*M* = 3.40, *SD* = 1,42, McDonald's ω = .88) was measured by five statements on the extent to which participants would like, comment on, or share a post (J. W. Kim, 2018), such as *this fact-checking post is worth sharing with others on social media* and *I would share this fact-checking post on my social media*.

AI familiarity. The experience and familiarity with AI (*Min* = 0, *Max* = 10, *M* = 6.44, *SD* = 2.81) were measured with one question on the knowledge of AI products – *which of the following technologies, if any, uses artificial intelligence (AI)?* (Wojcieszak et al., 2021). All ten possible answers were AI products, and the number of products selected indicated the familiarity with AI.

Conspiracy ideation. Conspiracy ideation (*M* = 2.77, *SD* = 0.82, McDonald's ω = .88) consisted of 10 items from a conspiracy ideation scale (Brotherton et al., 2013) on a 5-point scale (1 – disagree, 5 – agree). Sample statements are *the spread of certain viruses and/or diseases is the result of the deliberate, concealed efforts of some organization*, and *a lot of important information is deliberately concealed from the public out of self-interest*.

Prior attitude. Topic-specific prior attitude was measured by the extent to which one would agree with three topic-related statements on a 7-point scale (1 – disagree, 7 – agree), including *the 2020 presidential election was rigged* (*M* = 2.25, *SD* = 1.49), *the benefits of the COVID-19 vaccine outweigh its risks* (*M* = 5.89, *SD* = 1.58), and *the U.S. needs a rise in the minimum wage* (*M* = 5.85, *SD* = 1.38). These statements were selected from recent fact-checking messages and research papers (i.e., Zhang et al., 2021) to ensure timeliness and relevance.[1]

Prior attitude consistency. To answer RQ3 and RQ4, a binary variable was created to indicate whether a message was consistent with one's prior attitude on this issue. It was considered as consistent when participants agreed with the statements above on the Covid-19 vaccines and the rise of minimum wage, and when participants disagreed with the statement on the election fraud, and vice versa. Those who either agreed or disagreed were considered as

[1] It should be acknowledged that the statement regarding the federal minimum wage is purely an attitude statement instead of a true/false statement, which is essentially distinctive from the other two statements.

having no clear attitude on an issue, and therefore were removed from this analysis. The stimuli message on the Covid-19 vaccines was counter-attitudinal for 66 participants (10.7%; $N_{pro-attitudinal}$ = 548; $N_{no-attitude}$ = 43); the message on the election fraud was counter-attitudinal for 161 participants (10.3%; $N_{pro-attitudinal}$ = 495; $N_{no-attitude}$ = 104); the message on the minimum wage was counter-attitudinal for 49 participants (8.1%; $N_{pro-attitudinal}$ = 557; $N_{no-attitude}$ = 50).

Party affiliation. Party affiliation was measured with one item with responses ranging from a strong democrat to a strong republican (see Table 4 for a full list). This variable was quantified to a continuous variable from 1 (democrat) to 7 (republican) for further analysis.

Political ideology. Participants were asked to rate how left or right they are on a 11-point scale from 1 (political left) to 11 (political right). The variable was leaning toward left (M = 4.10, SD = 2.14).

Demographics. Participants were asked to report their age, gender, education, race, and income. For analysis, gender and race were converted to binary variables (gender: 1 – female, 0 – male; race: 1 – white, 0 – non-white); education and income were converted to continuous variables.

Results

Descriptive statistics. The mean differences of three dependent variables across language intensity, fact-checking agency, and topic were presented in Figure 1 and 3. The mean differences across language intensity and the agency were extremely close and non-significant, though dependent variables with high language intensity tended to be lower than the other conditions. It suggested that the differences in language intensity of fact-checking labels did not transfer to the intensity of the whole message. Besides, the Covid-19 vaccines was the topic with highest message intensity, message credibility, and engagement intentions; the topic was a significant factor that influenced dependent variables (F = 27.91, p < .001).

Analysis. Since this study involves three dependent variables – message intensity, message credibility, and engagement intentions, multivariate analyses of variance (MANOVA) were performed at first to test for a significant difference of means across conditions. To answer all RQs and test H1, linear effects models were conducted in R with the *lmerTest* package (Kuznetsova et al., 2017). The random order that three stimuli were presented to each participant were included as a random effect, in addition to the random intercept for participants. Language intensity and agency were included as fixed factors. To answer H2, multiple linear regressions were conducted with the agency evaluation as the dependent variable. For all regression models conducted, p-values reported were not adjusted for multiple comparisons. This was based on the fact that adjustment methods do not increase the disjunctive power or reduce the Type I error significantly when there are a few missing values only and the correlation between dependent variables is moderate (r = .42; Vickerstaff et al., 2019). Therefore, adjustments were not made at a risk of increasing a Type II error.

MANOVA tests were conducted on all messages that participants were exposed to (N = 1,971) with message intensity, message credibility, and engagement intentions as the dependent variables. It showed that the mean difference of three dependent variables was not significant cross three levels of language intensity ($F = 0.71$, $p = .640$) or across two different agencies ($F = 0.21$, $p = .893$).

RQ1 asked about the effect of language intensity. Linear effects models were conducted for each dependent variable (Table 5); the interaction with fact-checking agency was added in the model as well (Table 6). The results showed that language intensity was associated with lower levels of message credibility and engagement intentions. Specifically, high intensity negatively predicted engagement intentions, and mid intensity negatively predicted message credibility. High intensity was negatively associated with message credibility when the interaction was included in the model. Overall, language intensity was partially associated with lower levels of message credibility and engagement intentions.

H1 predicted that individuals would have better perceptions on fact-checking messages from AI than from humans, and RQ2 asked whether agency would affect engagement intentions. Linear regression models were run for each dependent variable respectively (Table 5, 6). Agency had no effect on any dependent variables; H1 was therefore rejected, and agency did not affect engagement intentions. H2 predicted that AI would be evaluated as a better fact-checking agency than humans. Two linear regressions were conducted, and there was no significant effect (Table 7). H2 was rejected as well. Overall, participants evaluated AI and human fact-checking agencies similarly.

Effects of motivated reasoning. RQ3 and RQ4 asked about the effect of language intensity and agency when it comes to counter-attitudinal fact-checking messages. Before answering these RQs, a MANOVA test showed that motivated reasoning had a significant effect on three dependent variables ($F = 15.55$, $p < .001$). Specifically, participants who found fact-checking messages counter-attitudinal tended to perceive the messages as less intense ($M_{counter} = 4.59$, $M_{pro} = 4.77$, $p = .013$) and less credible ($M_{counter} = 4.59$, $M_{pro} = 5.11$, $p < .001$), and they were less likely to engage with them ($M_{counter} = 3.07$, $M_{pro} = 3.45$, $p < .001$), compared with those who found messages as pro-attitudinal.

MANOVA tests of language intensity and agency showed that neither language intensity ($F = 0.19$, $p = .980$) nor agency ($F = 0.25$, $p = .860$) had a significant effect on these dependent variables. The results of linear regression models were reported in Table 8 (N = 172), and the results for pro-attitudinal messages (N = 1,600) and no-attitude messages (N = 197) were reported in Table 9 and 10 for reference. Still, language intensity or agency had no significant effect, though language intensity negatively predicted message credibility and engagement intentions when fact-checking messages were pro-attitudinal. For participants who held no obvious attitudes, language intensity positively predicted message credibility. Overall, the effect of language intensity was negligible when fact-checking messages were counter-attitudinal, while the effect of fact-checking agency remained null.

Effects of covariates. In addition to the effects of fixed factors, abovementioned full models also suggested significant effects explained by covariates. First, participants tended to perceive fact-checking messages as more intense and were more likely to engage when the messages were about the Covid-19 vaccines or election fraud, compared with the topic of minimum wage; participants tended to perceive fact-checking messages as more credible when the messages were about the Covid-19 vaccines than the other two topics (Table 5). The effect of the topic on engagement intentions existed when participants found the fact-checking messages pro-attitudinal (Table 9). Further, conspiracy ideation and political ideology were consistently associated with message credibility (Table 5-9); participants with higher levels of conspiracy ideation and right thinking tended to perceive fact-checking messages as less credible, regardless of one's prior attitude on the fact-checking subject. Lastly, female participants tended to perceive fact-checking messages as less credible, while white participants and those with higher income tended to perceive fact-checking messages as more credible (Table 5, 6). This demographic effect was especially true when the fact-checking messages were pro-attitudinal, though it was uncertain whether this was an effect of majority or of motivated reasoning, given that the pro-attitudinal population outnumbered the counter-attitudinal significantly.

Discussion

This study investigated the effect of language intensity and fact-checking agency with a factorial experiment. It revealed that language intensity negatively predicted message credibility; participants tended to perceive fact-checking messages as less credible when the messages employed intense language. This result provided an opposite evidence to the previous finding that language intensity was positively associated with persuasiveness and behavior change (Andersen & Blackburn, 2009; Buller et al., 2000). This might be explained by the difference in persuasion contexts. Positive effects of language intensity occurred in contexts like political speech and advocacy of health behaviors, where people expect intense language and strong argument, and high language intensity fits such expectation. According to the Language Expectancy Theory (Burgoon et al., 2002), high language intensity is a positive violation of language expectancy and therefore promotes persuasiveness. However, people are likely to expect fact-checking messages to be objective and neutral, and high language intensity may not be a positive violation of this expectation.

The second finding is that language intensity negatively indicated engagement intentions: people were less likely to engage with the messages when the messages employed intense language. This finding contradicted the observational study where language intensity was positively associated with engagement. This inconsistency might be explained by the heuristic-systematic processing model (HSM; Chaiken & Ledgerwood, 2012). Social media users were likely to engage with heuristic processing of fact-checking messages and therefore were likely to engage with intense messages, while participants in this experiment were asked to carefully read and evaluate the messages before answering questions on engagement intentions. The systematic processing of fact-checking messages may inhibit participants from engaging with the messages.

Regarding motivated reasoning, the effect of language intensity held true especially when it came to pro-attitudinal messages, while language intensity had no effect when the fact-checking messages were counter-attitudinal. It implies that language intensity did not matter when the fact-checking messages were against one's pre-existing attitudes, or that pre-existing attitudes tended to override the message effects. The null effect can also be explained by the small sample size of counter-attitudinal messages (N = 167); it is harder to obtain a significant effect when it comes to a small sample size. Besides, it is interesting that language intensity had a positive effect on message credibility when participants had no or neutral attitudes. When people have no obvious stance or have insufficient knowledge on an issue, people are likely to be more open-minded to the influence of fact-checking messages. This open-mindedness was confirmed in the Differential Information Model (Li & Wagner, 2020), which suggests that the uninformed people are more likely to experience belief updating after exposure of misinformation corrections.

The fact-checking agency had no significant effect on how people perceive fact-checking messages or evaluate fact-checking agencies. This finding was consistent with the argument of CASA that AI and humans are perceived as the same (Nass et al., 1994). In the case of fact-checking, people may perceive AI as just capable as professional journalists. This perception remained the same regardless of one's knowledge and familiarity with AI, given that AI knowledge had no significant effect (Table 6, 7). However, it does not suggest that people treat all kinds of fact-checking agencies as the same, since only AI and human independent agencies were studied here. Therefore, this finding only implies that people treat human and AI independent fact-checking agencies without difference, since a previous research has found that people trusted universities and health institutions more (Zhang et al., 2021).

Furthermore, the fact-checking topics, individual characteristics such as conspiracy ideation and political ideology, as well as demographics significantly influenced how people evaluate fact-checking messages and agencies. For example, participants tended to perceive fact-checking messages on the Covid-19 vaccines as more intense and more credible, and participants were more likely to engage with such messages. It suggested that topics with higher priority and recency tended to attract more intense responses. Besides, participants with higher levels of conspiracy ideation tended to perceive fact-checking messages as less credible and were less likely to engage. It makes sense since conspiracy ideation has been found to decrease science acceptance (Lewandowsky et al., 2013). Unsurprisingly, factors like topics and individual differences had larger effects than the manipulated factors in this study; it requires a more representative sample to establish a more solid conclusion.

The major limitation of this study was that language intensity was induced by one adjective only for each condition. It may bring in ambiguity to the results, such that the effect of language intensity might result from these particular adjectives instead of the intended language intensity. Further, a college-student sample was used in this study, and it might largely influence the findings, since this sample was extremely young, relatively wealthy, very liberal with low belief in misinformation, and familiar with AI technology. Therefore, it is necessary to replicate this study with a more representative sample.

Experiment 2

Together, the observational study and Experiment 1 has confirmed the partial negative effect of language intensity on message credibility. It is possible that the effect found in Experiment 1 resulted from the unique sample characteristics and would not be replicated with a more representative sample. Therefore, Experiment 2 was conducted exactly as Experiment 1 with a different sample recruited from Amazon Mechanical Turk. This experiment was meant to replicate the key findings regarding message credibility and engagement intentions.

Methods

Sample and procedure. American participants (N = 792) were recruited from MTurk with a compensation of $1. Participants were redirected to the same Qualtrics survey once they accepted the task (N = 960). The study design and procedure were kept constant; 5 non-adult participants and 1 without signing the consent were removed. Non-American responses (N = 150) and blank responses were removed as well (N = 12). Therefore, the final sample was reduced to N = 792 (517 male, 274 female). Participants were aged from 19 to 78 (M = 36.49, SD = 10.77) and well educated with 88% having a bachelor's degree or higher (N = 694). Participants' median annual household income fell in the $40,000 to $59,999 range. A detailed summary of sample characteristics is summarized in Table 4. Compared with the SONA sample, the MTurk sample was more balanced in terms of age, race, and income. The MTurk sample was more educated than the national average, and there were more males as well. There was no significant difference in age, gender, education, or income across conditions.

Results

Descriptive statistics. The mean differences of three dependent variables across language intensity, fact-checking agency, and topic were presented in Figure 2 and 4. Identical to Experiment 1, dependent variables did not vary significantly across language intensity or agency. However, MTurk participants tended to have significantly higher levels of message intensity ($M_{MTurk} = 5.23$, $M_{SONA} = 4.74$, $p < .001$), message credibility ($M_{MTurk} = 5.21$, $M_{SONA} = 5.04$, $p < .001$), and engagement intentions ($M_{MTurk} = 4.96$, $M_{SONA} = 3.40$, $p < .001$). Besides, dependent variables did not vary across topics ($F = 0.73$, $p = .623$). The MTurk sample was significantly different from the SONA sample in Experiment 1. The MTurk sample was less familiar with AI ($M_{MTurk} = 3.94$, $M_{SONA} = 6.44$, $p < .001$) and held more conspiracy ideation ($M_{MTurk} = 3.56$, $M_{SONA} = 2.77$, $p < .001$).

Similarly, MANOVA tested were conducted first to test if the mean differences of message credibility and engagement intentions vary across experiment conditions. The results suggested that the mean difference of three dependent variables was not significant cross three levels of language intensity ($N = 2,376$; $F = 0.75$, $p = .606$) or across two different agencies ($F = 1.14$, $p = .330$).

To answer RQ1 about the effect of language intensity, identical linear regression models were conducted. It showed that mid intensity negatively predicted message credibility; language intensity had no effect on message intensity or engagement intentions (Table 11). However, this effect diminished when the interaction with fact-checking agency was added to the model (Table 12). Overall, language intensity was partially associated with message credibility.

Similarly, linear effects model and a linear regression model were conducted to test H1 and H2 as well as to answer RQ2 on the effect of the fact-checking agency. Fact-checking agency had no significant effect on any of the dependent variables (Table 11,12). H2 predicted the relationship between fact-checking agency and agency evaluation. Linear regression models showed no effect (Table 13). Therefore, H1 and H2 were therefore rejected, and the agency did not affect engagement intentions.

Effects of motivated reasoning. Among all messages exposed to participants (N = 2,376), 620 were found to be counter-attitudinal, 1,473 were found to be pro-attitudinal, and 283 were found to be neutral to participants. Especially, the message on the election fraud was counter-attitudinal to 48.1% participants (N = 381). Identical to Experiment 2, motivated reasoning was a significant factor influencing dependent variables ($F = 15.06$, $p < .001$). Participants who found the fact-checking messages counter-attitudinal still had lower levels of message intensity ($M_{counter} = 5.10$, $M_{pro} = 5.29$, $p < .001$) and credibility ($M_{counter} = 4.98$, $M_{pro} = 5.30$, $p < .001$) and were less likely to engage with messages ($M_{counter} = 4.83$, $M_{pro} = 5.02$, $p < .001$).

MANOVA tests showed neither language intensity ($F = 0.63$, $p = .707$) nor fact-checking agency ($F = 1.15$, $p = .328$) influenced dependent variables when the messages were counter-attitudinal. The results of linear effects models of counter-attitudinal, pro-attitudinal, no-attitude messages were reported in Table 14 to 16, and no significant effect was found from language intensity or the fact-checking agency. Therefore, the negative effects of language intensity on message credibility or engagement intentions were not replicated in this experiment when the fact-checking messages were pro-attitudinal, while the null effects of fact-checking agency were replicated.

Effects of covariates. First, participants in Experiment 2 tended to care more about the issue of minimum wage than the Covid-19 vaccines or the election fraud: messages on the latter two topics were perceived as less intense and less credible, and participants were less likely to engage with these messages (Table 10). Besides, AI familiarity, conspiracy ideation and political ideology were surprisingly positively associated with agency evaluation, message credibility, and engagement intentions, regardless of motivated reasoning. It means that participants who tended to believe in conspiracy theories or were more politically right perceived fact-checking messages as more intense and more credible and were more likely to engage with these messages. Further, participants with stronger affiliation to the republican party tended to have lower levels of agency evaluation, message intensity, message credibility, and engagement intentions; the significant effect was less strong when the fact-checking messages were pro-attitudinal.

Lastly, female participants tended to perceive fact-checking messages as more intense and more credible, while white participants tended to perceive fact-checking messages as less intense and less credible. Participants with higher levels of education and income were more likely to engage with fact-checking messages. Similarly, the demographic effect was stronger when fact-checking messages were pro-attitudinal.

Discussion

Experiment 2 partially replicated and confirmed the key findings in Experiment 1 and the observational study. With a different sample, language intensity was again found to be a negative indicator of message credibility, though this effect was less strong, and there was no significant relationship between language intensity and engagement intentions. It might be because the MTurk sample was extremely engaging with the fact-checking messages, and this stable enthusiasm was not wavered by language intensity, fact-checking agency, or even topics in the messages. However, participants recruited through SONA were less engaging, but they paid more attention to the messages with high language intensity or about the Covid-19 vaccines. The discrepancy of the effect of engagement between the observational study and two experiments could also be explained by social desirability bias – it is possible that participants behave differently from in real life (Antin & Shaw, 2012). Overall, the negative effect of language intensity on message credibility was confirmed.

Consistently, the fact-checking agency did not affect the perceptions and evaluations of the fact-checking messages. With two different samples, the null effect of the fact-checking agency was confirmed, although this effect may only be applied to independent fact-checking agencies. These results were able to confirm the boundary condition of the machine heuristic, such that people perceive AI and human fact-checking agencies without difference. Besides, AI knowledge positively affected dependent variables; participants who were familiar with AI technology tended to find fact-checking messages more intense and credible and were more likely to engage with the messages.

Experiment 2 showed different effects of covariates on the dependent variables. First, a smaller effect from topics was observed. Participants viewed the messages on the minimum wage as more intense and more credible, though this effect did not transfer to engagement intentions. This might be explained by age and income differences – participants in the MTurk sample were older and less wealthy than those in the SONA sample, so it is possible that the MTurk sample cared more about the minimum wage than the SONA sample did.

Interestingly, conspiracy ideation showed a significant positive effect: participants with more conspiracy ideation tended to view fact-checking messages as more intense and more credible. There were similar opposite effects for gender, race and party affiliation; such differences were hard to explain, and it might result from differences in samples. It should be acknowledged that this sample was not common in the sense that the political ideology and political affiliation were not highly corrected (r = 0.16). Besides, there was a disproportionately large amount of participants with vaccine hesitancy and strong beliefs in misinformation such as the election fraud. This might be related to the low compensation of this study ($1 for each participant) as participants may not pay much attention to answer the questions. Overall, sample characteristics play a significant role in both experiments.

General discussion

Language intensity is a trivial linguistic feature that people tend to ignore but it can affect how we think implicitly. The current research extends the study of language intensity to the fact-checking context with a mix of an observational study and two factorial experiments. How people evaluate fact-checking messages and agencies and how language intensity and fact-checking agency (AI vs. human) affect such evaluations were examined. The most significant discovery of this research was that language intensity negatively predicted message credibility (Experiment 1 and 2). Besides, different effects of language intensity on engagement were found in the observational study and experiments; a null effect of fact-checking agency was confirmed in two experiments. Inconsistent effects from conspiracy ideation (Experiment 1 and 2), AI familiarity, party affiliation, political ideology, and demographics (i.e., gender, race, income) were also observed.

The current research contributed to the literature of language intensity by extending it to the fact-checking context. Fact-checking messages with intense adjectives elicited lower levels of message credibility. It suggested that people indeed pay attention to the variations in language intensity and value news objectivity, which affects how people evaluate content credibility and source objectivity. Belief updating can be harder if people do not trust fact-checking agencies.

However, inconsistent results were found regarding the effect on engagement. It might result from sample differences or social desirability bias in participants (Antin & Shaw, 2012); replication of this study is needed to confirm this effect. Besides, the effect of language intensity was negligible when the fact-checking messages were counter-attitudinal, which suggested that people tend not to trust the contents that are against their pre-existing attitudes. Though confirmation bias is hard to eliminate, language intensity still matters to fight against misinformation since credibility affects the effectiveness of misinformation corrections (Walter & Murphy, 2018). Therefore, this finding provides an important implication to fact-checking practitioners that fact-checking is more effective when neutral and unbiased language is employed.

Another important finding was about the effect of conspiracy ideation, AI familiarity, party affiliation, political ideology, and demographics. It was not surprising that these covariates possessed larger effects than the manipulated factors; people tend to behave in line with pre-existing attitudes and behaviors. A contradictory effect of conspiracy ideation and political ideology was found: negative effects were observed in Experiment 1 while positive effects were observed in Experiment 2. The negative effect aligned with previous findings that people with high conspiracy thinking tend to reject science, such as fact-checking (Lewandowsky et al., 2013). But it is hard to explain why participants with high conspiracy ideation found fact-checking messages more credible. Such inconsistency existed for AI familiarity and party affiliation, the effects of which were only found in Experiment 2. These results suggested the importance of samples; future research should replicate this study for these inconsistencies.

This study also contributed to the study of AI perception by identifying one boundary condition of machine heuristics (Sundar & Kim, 2019). Fact-checking is a context in which people treat AI and human agency without difference by testing with two different samples. This finding suggested that people perceive AI as capable as professional journalists in fact-checking, which is in line with the argument of CASA (Nass et al., 1994) and a recent finding in the context of online content moderation (Wojcieszak et al., 2021). This result also contributed by including AI as a potential fact-checking agency, since current research mainly focuses on existing fact-checking agencies such as news agency and universities. It implies that AI fact-checking could be promoted more broadly since it may achieve the same effect as humans do and greatly relieve the pressure from professional journalists.

Limitations and future research

There are a few limitations of this study. First, this study mainly focused on independent fact-checking agencies, while other institutions that also contribute to fact-checking, such as general news agencies and universities, were not studied here. The general public may not know independent fact-checking agencies as well as other familiar institutions, while independent fact-checking agencies may not be exposed to a representative audience. Second, in the experiments, only one adjective was employed to induce each level of language intensity. It hurt the internal validity of this study since the effect of language intensity may be induced by these specific adjectives rather than different levels of language intensity that they represent.

Though this study aimed to be as comprehensive as possible, several factors can be studied in the future research. First, given that a piece of information can be verified along the spectrum from true to false, it is worth studying both true and false fact-checking labels. Second, the target of fact-checking was not controlled in the experiment. Though the targets of fact-checking were mainly individuals with a high saying such as public figures and celebrities, random social media users could still post misinformation and become the targets as well. It would be interesting to test how people respond differently to fact-checking of public figures and random users. Third, more topics should be covered in the stimuli. This study only covered three different issues and significant effects have been observed. With more topics covered, we will be able to obtain a more reliable conclusion regarding the manipulated factors.

References

Abelson, R. P. (1995). Attitude extremity. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 25–42). Lawrence Erlbaum.

Andersen, P. A., & Blackburn, T. R. (2009). An experimental study of language intensity and response rate in e mail surveys. *Communication Reports*, *21*(1), 73–84. https://doi.org/10.1080/08934210409389377

Antin, J., & Shaw, A. (2012). Social desirability bias and self-reports of motivation: A study of Amazon Mechanical Turk in the US and India. *Conference on Human Factors in Computing Systems - Proceedings*, 2925–2934. https://doi.org/10.1145/2207676.2208699

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY*, *35*(3), 611–623. https://doi.org/10.1007/S00146-019-00931-W

Athanasiadou, A. (2007). On the subjectivity of intensifiers. *Language Sciences*, *29*(4), 554–565. https://doi.org/10.1016/J.LANGSCI.2007.01.009

Bode, L., & Vraga, E. K. (2017). See Something, Say Something: Correction of Global Health Misinformation on Social Media. *Health Communication*, *33*(9), 1131–1140. https://doi.org/10.1080/10410236.2017.1331312

Bowers, J. W. (1963). Language intensity, social introversion, and attitude change. *Communication Monographs*, *30*(4), 345–352. https://doi.org/10.1080/03637756309375380

Brandtzaeg, P. B., Følstad, A., & Domínguez, M. Á. C. (2017). How Journalists and Social Media Users Perceive Online Fact-Checking and Verification Services. *Journalism Practice*, *12*(9), 1109–1129. https://doi.org/10.1080/17512786.2017.1363657

Brave, S., Nass, C., & Hutchinson, K. (2005). Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-Computer Studies*, *62*(2), 161–178. https://doi.org/10.1016/J.IJHCS.2004.11.002

Brotherton, R., French, C. C., & Pickering, A. D. (2013). Measuring Belief in Conspiracy Theories: The Generic Conspiracist Beliefs Scale. *Frontiers in Psychology*, *4*(MAY), 279. https://doi.org/10.3389/fpsyg.2013.00279

Buller, D. B., Burgoon, M., Hall, J. R., Levine, N., Taylor, A. M., Beach, B., Buller, M. K., & Melcher, C. (2009). Long-Term Effects of Language Intensity in Preventive Messages on Planned Family Solar Protection. *Health Communication*, *12*(3), 261–275. https://doi.org/10.1207/S15327027HC1203_03

Buller, D. B., Burgoon, M., Hall, J. R., Levine, N., Taylor, A. M., Beach, B. H., Melcher, C., Buller, M. K., Bowen, S. L., Hunsaker, F. G., & Bergen, A. (2000). Using Language Intensity to Increase the Success of a Family Intervention to Protect Children from Ultraviolet Radiation: Predictions from Language Expectancy Theory. *Preventive Medicine*, *30*(2), 103–113. https://doi.org/10.1006/PMED.1999.0600

Burgers, C., & de Graaf, A. (2013). Language intensity as a sensationalistic news feature: The influence of style on sensationalism perceptions and effects. *Communications*, *38*(2), 167–188. https://doi.org/10.1515/commun-2013-0010

Burgoon, M., Denning, V. P., & Roberts, L. (2002). Language expectancy theory. In J. P. Dillard & M. Pfau (Eds.), *The persuasion handbook: Developments in theory and practice* (pp. 117–133). Sage Publications.

Chaiken, S., & Ledgerwood, A. (2012). A theory of heuristic and systematic information processing. *Handbook of Theories of Social Psychology: Volume 1*, 246–266. https://doi.org/10.4135/9781446249215.N13

Clementson, D. E., Pascual-Ferrá, P., & Beatty, M. J. (2015). How Language Can Influence Political Marketing Strategy and a Candidate's Image: Effect of Presidential Candidates' Language Intensity and Experience on College Students' Ratings of Source Credibility. *Journal of Political Marketing*, *15*(4), 388–415. https://doi.org/10.1080/15377857.2014.959689

Conger, K. (2020, September 15). *Driver charged in Uber's fatal 2018 autonomous car crash*. The New York Times.

Craig, T. Y., & Blankenship, K. L. (2011). Language and Persuasion: Linguistic Extremity Influences Message Processing and Behavioral Intentions: *Journal of Language and Social Psychology*, *30*(3), 290–310. https://doi.org/10.1177/0261927X11407167

CrowdTangle Team. (2021). *CrowdTangle*. Facebook.

Dastin, J. (2018, October 10). *Amazon scraps secret AI recruiting tool that showed bias against women | Reuters*. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2014). Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2466040

Druckman, J. N., & Bolsen, T. (2011). Framing, motivated reasoning, and opinions about emergent technologies. *Journal of Communication*, *61*(4), 659–688. https://doi.org/10.1111/j.1460-2466.2011.01562.x

Eagly, A. H., Wood, W., & Chaiken, S. (1978). Causal inferences about communicators and their effect on opinion change. *Journal of Personality and Social Psychology*, *36*(4), 424–435.

https://doi.org/10.1037/0022-3514.36.4.424

Gambino, A., Fox, J., & Ratan, R. (2020). Building a Stronger CASA: Extending the Computers Are Social Actors Paradigm. *Human-Machine Communication*, *1*(1), 5. https://doi.org/10.30658/hmc.1.5

Ginsberg, M. (2012). Introduction: What is AI? In *Essentials of Artificial Intelligence* (pp. 3–17). Elsevier Science.

Graefe, A., & Bohlken, N. (2020). Automated Journalism: A Meta-Analysis of Readers' Perceptions of Human-Written in Comparison to Automated News. *Media and Communication*, *8*(3), 50–59. https://doi.org/10.17645/MAC.V8I3.3019

Grant, M. (2000). *Greek and roman historians: information and misinformation*. Routledge.

Graβl, P., Schaap, G., Spagnuolo, F., & Riet, J. V. 't. (2019). The effects of scandalization in political news messages on political trust and message evaluation: *Journalism*. https://doi.org/10.1177/1464884919879582

Hamilton, M. A. (1998). Message Variables That Mediate and Moderate the Effect of Equivocal Language on Source Credibility: *Journal of Language and Social Psychology*, *17*(1), 109–143. https://doi.org/10.1177/0261927X980171006

Hamilton, M. A., Hunter, J. E., & Burgoon, M. (1990). An Empirical Test of an Axiomatic Model of the Relationship Between Language Intensity and Persuasion. *Journal of Language and Social Psychology*, *9*(4), 235–255. https://doi.org/10.1177/0261927X9094002

Hamilton, M. A., & Stewart, B. L. (1993). Extending an information processing model of language intensity effects. *Communication Quarterly*, *41*(2), 231–246. https://doi.org/10.1080/01463379309369882

Hart, P. S., & Nisbet, E. C. (2012). Boomerang Effects in Science Communication: How

Motivated Reasoning and Identity Cues Amplify Opinion Polarization About Climate Mitigation Policies. *Communication Research*, *39*(6), 701–723. https://doi.org/10.1177/0093650211416646

Hassan, N., Arslan, F., Li, C., & Tremayne, M. (2017). Toward automated fact-checking: Detecting check-worthy factual claims by claimbuster. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *Part F129685*, 1803–1812. https://doi.org/10.1145/3097983.3098131

Hong, J.-W., Choi, S., & Williams, D. (2020). Sexist AI: An Experiment Integrating CASA and ELM. *International Journal of Human–Computer Interaction*, *36*(20), 1928–1941. https://doi.org/10.1080/10447318.2020.1801226

Hong, J. W. (2020). Why Is Artificial Intelligence Blamed More? Analysis of Faulting Artificial Intelligence for Self-Driving Car Accidents in Experimental Settings. *International Journal of Human–Computer Interaction*, *36*(18), 1768–1774. https://doi.org/10.1080/10447318.2020.1785693

Hosman, L. A. (2002). Language and Persuasion. In J. P. Dillard & M. Pfau (Eds.), *The persuasion handbook developments in theory and practice* (pp. 371–390). Sage.

IBM Watson. (2021). *IBM Watson*.

Jünger, J., & Keyling, T. (2019). *Facepager. An application for automated data retrieval on the web.*

Karadzhov, G., Nakov, P., Barrón-Cedeño, A., & Koychev, I. (2017). Fully Automated Fact Checking Using External Sources. *RANLP 2017 - Recent Advances in Natural Language Processing Meet Deep Learning*, 344–353. https://doi.org/10.26615/978-954-452-049-6_046

Kim, C., & Yang, S. U. (2017). Like, comment, and share on Facebook: How each behavior differs from the other. *Public Relations Review*, *43*(2), 441–449. https://doi.org/10.1016/J.PUBREV.2017.02.006

Kim, J. W. (2018). Rumor has it: The effects of virality metrics on rumor believability and transmission on Twitter: *New Media & Society*, *20*(12), 4807–4825. https://doi.org/10.1177/1461444818784945

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(1), 1–26. https://doi.org/10.18637/JSS.V082.I13

Lewandowsky, S., Gignac, G. E., & Oberauer, K. (2013). The Role of Conspiracist Ideation and Worldviews in Predicting Rejection of Science. *PLOS ONE*, *8*(10), e75637. https://doi.org/10.1371/JOURNAL.PONE.0075637

Li, J., & Wagner, M. W. (2020). The Value of Not Knowing: Partisan Cue-Taking and Belief Updating of the Uninformed, the Ambiguous, and the Misinformed. *Journal of Communication*, *70*, 646–669. https://doi.org/10.1093/joc/jqaa022

Liebrecht, C., Hustinx, L., & Mulken, M. van. (2019). The Relative Power of Negativity: The Influence of Language Intensity on Perceived Strength: *Journal of Language and Social Psychology*, *38*(2), 170–193. https://doi.org/10.1177/0261927X18808562

Logg, J. M., Minson, J. A., & Moore, D. A. (2018). *Algorithm Appreciation: People Prefer Algorithmic To Human Judgment*.

Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (1998). Automation Bias: Decision Making and Performance in High-Tech Cockpits. *The International Journal of Aviation Psychology*, *8*(1), 47–63. https://doi.org/10.1207/S15327108IJAP0801_3

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computer are social actors. *Conference on Human Factors in Computing Systems - Proceedings*, 72–78. https://doi.org/10.1145/259963.260288

Oeldorf-Hirsch, A., Schmierbach, M., Appelman, A., & Boyle, M. P. (2020). The Ineffectiveness of Fact-Checking Labels on News Memes and Articles. *Mass Communication and Society*, *23*(5), 682–704. https://doi.org/10.1080/15205436.2020.1733613

Osgood, C. E. (1969). On the whys and wherefores of E, P, and A. *Journal of Personality and Social Psychology*, *12*(3), 194–199. https://doi.org/10.1037/H0027715

Shank, D. B., DeSanti, A., & Maninger, T. (2019). When are artificial intelligence versus human agents faulted for wrongdoing? Moral attributions after individual and joint decisions. *Information, Communication & Society*, *22*(5), 648–663. https://doi.org/10.1080/1369118X.2019.1568515

Shin, D. (2021). The perception of humanness in conversational journalism: An algorithmic information-processing perspective. *New Media & Society*, 146144482199380. https://doi.org/10.1177/1461444821993801

Sundar, S. (2008). The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 73–100). The MIT Press. https://doi.org/10.1162/dmal.9780262562324.073

Sundar, S. S., & Kim, J. (2019). Machine Heuristic: When We Trust Computers More Than Humans With Our Personal Information. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3290605

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*(3), 755–769. https://doi.org/10.1111/j.1540-5907.2006.00214.x

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

Vickerstaff, V., Omar, R. Z., & Ambler, G. (2019). Methods to adjust for multiple comparisons in the analysis and sample size calculation of randomised controlled trials with multiple primary outcomes. *BMC Medical Research Methodology*, *19*. https://doi.org/10.1186/s12874-019-0754-4

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., & 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

Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1–37. https://doi.org/10.2478/JAGI-2019-0002

Wojcieszak, M., Thakur, A., Ferreira Gonçalves, J. F., Casas, A., Menchen-Trevino, E., & Boon, & M. (2021). Can AI Enhance People's Support for Online Moderation and Their Openness to Dissimilar Political Views? *Journal of Computer-Mediated Communication*. https://doi.org/10.1093/JCMC/ZMAB006

Zarouali, B., Makhortykh, M., Bastian, M., & Araujo, T. (2020). Overcoming polarization with chatbot news? Investigating the impact of news content containing opposing views on agreement and credibility: *European Journal of Communication*, *36*(1), 53–68. https://doi.org/10.1177/0267323120940908

Zhang, J., Featherstone, J. D., Calabrese, C., & Wojcieszak, M. (2021). Effects of fact-checking social media vaccine misinformation on attitudes toward vaccines. *Preventive Medicine*, *145*, 106408. https://doi.org/10.1016/j.ypmed.2020.106408

Table 1. Descriptive statistics of 30 veracity-related adjectives.

adjective	frequency in Facebook posts	Pilot study of language intensity							
		N	Min	Max	M	SD	Median	skewness	kurtosis
manipulated	27	55	2.0	7.0	5.60	1.18	5.80	-0.61	-0.27
fraudulent	31	61	2.4	7.0	5.35	1.04	5.40	-0.61	0.30
deceptive	15	64	2.0	7.0	5.12	1.04	5.25	-0.51	0.64
inflammatory	41	64	2.4	7.0	5.12	1.18	5.00	-0.03	-0.79
suspicious	28	65	2.0	7.0	4.84	1.03	4.90	0.06	-0.11
exaggerated	67	60	1.6	7.0	4.81	1.13	4.95	-0.63	0.76
fabricated	71	61	2.0	7.0	4.78	1.17	4.80	-0.02	-0.29
erroneous	41	61	1.0	7.0	4.77	1.23	4.70	-0.08	0.40
bizarre	56	62	1.5	7.0	4.67	1.21	4.65	-0.26	0.04
fake	762	66	1.0	7.0	4.65	1.33	4.70	-0.42	0.11
fictitious	29	62	1.0	7.0	4.49	1.26	4.45	-0.41	0.57
wrong	604	67	1.0	7.0	4.48	1.38	4.60	-0.44	0.39
distorted	24	65	2.3	7.0	4.45	0.92	4.40	-0.01	0.09
flawed	17	65	1.2	7.0	4.41	1.06	4.40	-0.44	0.77
misleading	853	67	1.3	7.0	4.32	1.16	4.40	-0.15	0.43
false	4047	64	1.1	7.0	4.19	1.34	4.00	-0.12	-0.38
inaccurate	255	65	1.5	7.0	4.18	1.09	4.20	0.03	0.21
incorrect	125	64	1.0	7.0	4.16	1.35	4.05	0.12	-0.41
invalid	6	67	1.0	7.0	4.15	1.29	4.30	0.00	-0.20
unsubstantiated	93	62	1.0	6.7	4.08	1.13	4.10	-0.54	0.60
phony	21	69	1.0	7.0	4.04	1.35	4.10	0.11	-0.53
dubious	143	64	1.7	7.0	3.87	1.14	3.70	0.49	0.27
untrue	41	67	1.0	6.1	3.80	1.33	4.10	-0.39	-0.91
misquoted	13	66	1.0	7.0	3.70	1.30	3.80	0.27	-0.10
unverified	27	64	1.0	6.0	3.57	1.12	3.70	-0.14	-0.23
fictional	121	61	1.0	5.8	3.52	1.19	3.60	-0.18	-0.88
unfounded	107	62	1.0	5.6	3.49	1.23	3.65	-0.41	-0.72
unproven	33	61	1.0	6.0	3.48	1.26	3.50	-0.07	-0.73
incomplete	66	61	1.0	6.2	3.41	1.27	3.80	-0.35	-0.80
unclear	88	65	1.0	6.4	2.93	1.19	2.90	0.61	0.23

Table 2. Summary statistics of language intensity and engagement in fact-checking agencies.

Agency	# posts	#Comments	Intensity (post)		Emotionality (post)		Emotionality (comment)		#Engagement	
			M	SD	M	SD	M	SD	M	SD
AFP Fact Check	3107	2728	3.93	2.37	0.61	0.25	0.59	0.33	11.91	45.18
Fact Checker	2163	52372	0.82	2.06	0.55	0.35	0.58	0.17	165.39	309.29
FactCheck.org	2275	1278734	2.82	3.23	0.62	0.27	0.59	0.10	3777.18	4370.40
Full Fact	3226	194430	0.56	1.65	0.51	0.31	0.57	0.19	237.97	510.88
Health Feedback	32	255	2.99	4.03	0.52	0.17	0.58	0.33	72.81	149.31
PolitiFact	3016	909025	0.91	1.96	0.53	0.22	0.58	0.12	2273.67	3042.99
snopes.com	19936	1694754	0.71	1.81	0.53	0.35	0.56	0.19	433.44	1160.22
Sum	33755	4132298	1.16	2.26	0.54	0.33	0.57	0.18	748.22	1980.40

Table 3. Poisson regression models of language intensity, emotionality, and engagement.

	Engagement				Emotionality			
	b	95% CI		p	b	95% CI		p
		Lower	Upper			Lower	Upper	
Intercept	1.83	1.58	2.08	0.000	-0.60	-0.62	-0.59	0.000
Intensity	0.08	0.07	0.09	0.000	0.01	0.01	0.01	0.000
word count	0.01	0.01	0.01	0.000	0.00	0.00	0.00	0.000
#followers	0.00	0.00	0.00	0.000	0.00	0.00	0.00	0.199
Emotionality (post)	-0.12	-0.19	-0.04	0.003	0.04	0.03	0.06	0.000

Table 4. Demographic characteristics of the SONA and Mturk samples.

Characteristics	SONA sample		Mturk sample	
	Frequency	Percent	Frequency	Percent
Male	132	20.1	517	65.3
Age, mean (SD)	20.3 (2.85)	-	36.5 (10.77)	-
Race - Asian	277	42.2	51	6.4
Race - White	191	29.1	547	69.1
Race - Black	36	5.5	116	14.6
Race - Hispanic	184	28.0	80	10.1
Edu - H.S. /GED	190	28.9	21	2.7
Edu - Associate's degree	73	11.1	35	4.4
Edu - Some college no degree	333	50.7	37	4.7
Edu - Bachelor's degree	54	8.2	484	61.1
Edu - Graduate degree	1	0.2	210	26.5
Income - <$20,000	127	19.3	58	7.3
Income - $20,000–$39,999	86	13.1	149	18.8
Income - $40,000–$59,999	51	7.8	277	35.0
Income - $60,000–$79,999	43	6.5	141	17.8
Income - $80,000–$99,999	32	4.9	109	13.8
Income - $100,000–$149,999	82	12.5	47	5.9
Income - ≥$150,000	88	13.4	8	1.0
Income - Prefer not to say	147	22.4	3	0.4
A strong Democrat	183	27.9	348	43.9
A not very strong Democrat	166	25.3	93	11.7
Independent, lean toward Democrat	166	25.3	60	7.6
Independent (close to neither party)	99	15.1	59	7.4
Independent, lean toward Republican	28	4.3	59	7.4
A not very strong Republican	6	0.9	62	7.8
A strong Republican	8	1.2	111	14.0
Left-Right, mean (SD)	4.1 (2.14)	-	8.04 (2.86)	-
Native speaker of English	490	74.58	787	99.37

Table 5. Linear effects models: the effect of language intensity and fact-checking agency on message intensity, message credibility, and engagement intentions (SONA).

	Message intensity				Message credibility				Engagement			
	b	95% CI		p	b	95% CI		p	b	95% CI		p
		Lower	Upper			Lower	Upper			Lower	Upper	
Intercept	4.37	3.69	5.05	0.000	6.09	5.35	6.83	0.000	3.77	2.87	4.67	0.000
Intensity.Mid	0.01	-0.08	0.10	0.840	-0.10	-0.20	0.00	0.046*	-0.08	-0.19	0.03	0.169
Intensity.High	-0.01	-0.10	0.08	0.789	-0.07	-0.17	0.03	0.159	-0.14	-0.25	-0.03	0.014*
Agency.Human	0.03	-0.12	0.18	0.694	0.05	-0.11	0.22	0.520	0.07	-0.13	0.27	0.506
Topic.Vaccine	0.29	0.20	0.38	0.000***	0.24	0.14	0.33	0.000***	0.86	0.74	0.97	0.000***
Topic.Election	0.09	0.00	0.19	0.044*	-0.15	-0.25	-0.06	0.002**	0.29	0.18	0.40	0.000***
AI familiarity	0.02	-0.01	0.04	0.210	0.03	0.00	0.06	0.064	0.01	-0.02	0.05	0.513
Conspiracy ideation	-0.02	-0.12	0.08	0.658	-0.25	-0.36	-0.15	0.000***	-0.08	-0.21	0.05	0.226
Party affiliation	-0.03	-0.10	0.04	0.381	-0.02	-0.10	0.06	0.584	-0.08	-0.18	0.01	0.096.
Political ideology	-0.03	-0.07	0.02	0.258	-0.08	-0.13	-0.04	0.000***	-0.04	-0.10	0.01	0.151
Gender.Female	-0.01	-0.20	0.17	0.897	-0.22	-0.42	-0.02	0.031*	-0.10	-0.34	0.15	0.444
Age	0.02	-0.01	0.05	0.174	-0.01	-0.04	0.02	0.384	-0.01	-0.05	0.02	0.491
Education	-0.07	-0.17	0.03	0.162	0.08	-0.03	0.18	0.154	0.07	-0.06	0.20	0.294
Race.White	0.27	0.08	0.47	0.007**	0.24	0.04	0.45	0.023*	-0.13	-0.39	0.13	0.329
Income	0.02	0.00	0.04	0.032*	0.00	-0.02	0.02	0.658	0.00	-0.03	0.02	0.948

Table 6. Linear effects models: the effect of language intensity and fact-checking agency on message intensity, message credibility, and engagement intentions (interaction added; SONA).

	Message intensity				Message credibility				Engagement			
	b	95% CI Lower	95% CI Upper	p	b	95% CI Lower	95% CI Upper	p	b	95% CI Lower	95% CI Upper	p
Intercept	4.38	3.70	5.06	0.000	6.14	5.39	6.88	0.000	3.79	2.89	4.69	0.000
Intensity.Mid	0.01	-0.12	0.14	0.880	-0.15	-0.29	-0.01	0.037*	-0.11	-0.27	0.05	0.174
Intensity.High	-0.03	-0.16	0.10	0.632	-0.16	-0.30	-0.02	0.022*	-0.16	-0.32	0.00	0.051.
Agency.Human	0.02	-0.17	0.20	0.848	-0.04	-0.24	0.16	0.705	0.04	-0.20	0.27	0.771
Intensity.Mid:Human	0.00	-0.18	0.18	0.989	0.10	-0.10	0.29	0.329	0.06	-0.16	0.28	0.577
Intensity.High:Human	0.04	-0.14	0.22	0.682	0.18	-0.01	0.37	0.069	0.04	-0.18	0.26	0.749
Topic.Vaccine	0.29	0.20	0.38	0.000***	0.24	0.14	0.33	0.000***	0.85	0.74	0.97	0.000***
Topic.Election	0.09	0.00	0.19	0.046*	-0.16	-0.25	-0.06	0.002**	0.29	0.18	0.40	0.000***
AI familiarity	0.02	-0.01	0.04	0.210	0.03	0.00	0.06	0.064	0.01	-0.02	0.05	0.513
Conspiracy ideation	-0.02	-0.12	0.08	0.658	-0.25	-0.36	-0.15	0.000***	-0.08	-0.21	0.05	0.226
Party affiliation	-0.03	-0.10	0.04	0.381	-0.02	-0.10	0.06	0.584	-0.08	-0.18	0.01	0.096
Political ideology	-0.03	-0.07	0.02	0.258	-0.08	-0.13	-0.04	0.000***	-0.04	-0.10	0.01	0.151
Gender.Female	-0.01	-0.20	0.17	0.897	-0.22	-0.42	-0.02	0.031*	-0.10	-0.34	0.15	0.444
Age	0.02	-0.01	0.05	0.174	-0.01	-0.04	0.02	0.384	-0.01	-0.05	0.02	0.491
Education	-0.07	-0.17	0.03	0.162	0.08	-0.03	0.18	0.154	0.07	-0.06	0.20	0.293
Race.White	0.27	0.08	0.47	0.007**	0.24	0.04	0.45	0.023*	-0.13	-0.39	0.13	0.329
Income	0.02	0.00	0.04	0.032*	0.00	-0.02	0.02	0.658	0.00	-0.03	0.02	0.948

Table 7. Linear regression models: the effect of fact-checking agency on agency evaluation (SONA).

	b	95% CI		p
		Lower	Upper	
Intercept	5.43	4.53	6.33	0.000
Agency.Human	0.04	-0.17	0.25	0.679
AI knowledge	0.03	-0.01	0.06	0.173
Conspiracy ideation	-0.28	-0.42	-0.15	0.000***
Party affiliation	-0.04	-0.15	0.05	0.307
Political ideology	-0.04	-0.11	0.01	0.106
Gender.Female	0.21	-0.05	0.46	0.113
Age	0.01	-0.03	0.05	0.614
Education	0.03	-0.11	0.16	0.670
Race.White	0.13	-0.14	0.40	0.347
Income	0.01	-0.01	0.04	0.287

Table 8. Linear effects models: the effect of language intensity and fact-checking agency on message intensity, message credibility, and engagement intentions (counter-attitudinal posts; SONA).

	Message intensity				Message credibility				Engagement			
	b	95% CI		p	b	95% CI		p	b	95% CI		p
		Lower	Upper			Lower	Upper			Lower	Upper	
Intercept	2.54	0.46	4.54	0.022	5.24	3.29	7.08	0.000	3.80	1.56	6.07	0.002
Intensity.Mid	0.00	-0.41	0.40	0.993	0.02	-0.39	0.44	0.921	-0.24	-0.75	0.25	0.366
Intensity.High	-0.34	-0.73	0.07	0.108	-0.31	-0.69	0.13	0.149	-0.16	-0.64	0.31	0.527
Agency.Human	0.13	-0.30	0.55	0.575	0.05	-0.34	0.43	0.798	0.18	-0.29	0.64	0.489
Topic.Vaccine	0.37	-0.03	0.80	0.090	-0.08	-0.51	0.34	0.703	0.87	0.38	1.38	0.002**
Topic.Election	0.42	-0.01	0.84	0.064	-0.29	-0.71	0.14	0.207	0.44	-0.07	0.95	0.112
AI familiarity	0.06	-0.01	0.14	0.111	0.01	-0.06	0.08	0.777	-0.01	-0.09	0.07	0.869
Conspiracy ideation	-0.07	-0.34	0.20	0.620	-0.37	-0.61	-0.11	0.007**	-0.05	-0.36	0.25	0.747
Party affiliation	0.03	-0.13	0.20	0.731	0.08	-0.07	0.23	0.349	0.08	-0.11	0.26	0.448
Political ideology	0.01	-0.11	0.11	0.932	-0.19	-0.29	-0.09	0.001*	-0.06	-0.18	0.07	0.392
Gender.Female	-0.16	-0.63	0.31	0.525	-0.30	-0.72	0.13	0.193	-0.39	-0.90	0.13	0.175
Age	0.10	0.03	0.17	0.014*	0.06	-0.01	0.12	0.120	0.02	-0.06	0.10	0.665
Education	-0.19	-0.44	0.07	0.175	0.16	-0.07	0.39	0.204	-0.28	-0.56	0.01	0.072.
Race.White	0.27	-0.33	0.88	0.402	0.20	-0.34	0.75	0.492	-0.73	-1.40	-0.05	0.049*
Income	0.02	-0.04	0.07	0.537	0.00	-0.05	0.05	0.890	-0.01	-0.07	0.05	0.782

Table 9. Linear effects models: the effect of language intensity and fact-checking agency on message intensity, message credibility, and engagement intentions (pro-attitudinal posts; SONA).

	Message intensity				Message credibility				Engagement			
	b	95% CI		p	b	95% CI		p	b	95% CI		p
		Lower	Upper			Lower	Upper			Lower	Upper	
Intercept	4.49	3.78	5.20	0.000	6.18	5.39	6.97	0.000	3.58	2.65	4.51	0.000
Intensity.Mid	0.01	-0.09	0.11	0.852	-0.09	-0.20	0.01	0.086.	-0.09	-0.22	0.04	0.157
Intensity.High	0.02	-0.08	0.12	0.696	-0.07	-0.18	0.03	0.181	-0.14	-0.27	-0.02	0.026*
Agency.Human	0.02	-0.14	0.17	0.846	0.05	-0.13	0.22	0.605	0.05	-0.16	0.26	0.635
Topic.Vaccine	0.31	0.21	0.41	0.000***	0.27	0.17	0.38	0.000***	0.92	0.79	1.04	0.000***
Topic.Election	0.09	-0.01	0.20	0.075.	-0.14	-0.25	-0.03	0.013*	0.32	0.19	0.44	0.000***
AI familiarity	0.02	-0.01	0.04	0.282	0.03	0.00	0.06	0.042*	0.02	-0.02	0.05	0.438
Conspiracy ideation	-0.03	-0.13	0.08	0.638	-0.27	-0.38	-0.16	0.000***	-0.06	-0.20	0.07	0.355
Party affiliation	-0.01	-0.09	0.07	0.789	-0.01	-0.10	0.07	0.788	-0.07	-0.17	0.03	0.200
Political ideology	-0.03	-0.08	0.01	0.148	-0.08	-0.13	-0.03	0.003**	-0.04	-0.10	0.02	0.242
Gender.Female	-0.01	-0.20	0.19	0.957	-0.22	-0.44	-0.01	0.043*	-0.04	-0.29	0.22	0.778
Age	0.01	-0.02	0.04	0.507	-0.02	-0.05	0.01	0.268	-0.01	-0.05	0.02	0.472
Education	-0.04	-0.14	0.06	0.466	0.06	-0.05	0.17	0.322	0.08	-0.05	0.22	0.221
Race.White	0.26	0.06	0.47	0.012*	0.27	0.05	0.49	0.018*	-0.08	-0.35	0.19	0.573
Income	0.02	0.00	0.04	0.047*	0.01	-0.01	0.03	0.497	0.00	-0.02	0.03	0.946

Table 10. Linear effects models: the effect of language intensity and fact-checking agency on message intensity, message credibility, and engagement intentions (no-attitude posts; SONA).

	Message intensity				Message credibility				Engagement			
	b	95% CI		p	b	95% CI		p	b	95% CI		p
		Lower	Upper			Lower	Upper			Lower	Upper	
Intercept	1.55	-0.48	3.49	0.146	3.76	1.45	6.07	0.003	4.23	1.40	7.05	0.006
Intensity.Mid	0.21	-0.12	0.59	0.238	0.24	-0.21	0.68	0.317	0.15	-0.28	0.60	0.515
Intensity.High	0.49	0.13	0.88	0.014*	0.60	0.12	1.06	0.019*	0.40	-0.11	0.89	0.138
Agency.Human	0.11	-0.23	0.47	0.552	0.10	-0.29	0.49	0.636	0.01	-0.47	0.50	0.960
Topic.Vaccine	0.19	-0.24	0.61	0.397	0.01	-0.59	0.57	0.965	0.30	-0.31	0.85	0.317
Topic.Election	0.15	-0.22	0.55	0.446	-0.06	-0.56	0.41	0.815	0.55	0.04	1.05	0.040*
AI familiarity	-0.02	-0.08	0.04	0.604	-0.01	-0.09	0.06	0.704	0.06	-0.03	0.15	0.241
Conspiracy ideation	0.01	-0.26	0.25	0.950	-0.17	-0.47	0.12	0.271	-0.01	-0.37	0.35	0.965
Party affiliation	-0.28	-0.42	-0.13	0.001**	-0.17	-0.33	0.00	0.072.	-0.19	-0.40	0.02	0.094.
Political ideology	0.13	0.04	0.23	0.012**	0.10	-0.02	0.21	0.116	0.07	-0.07	0.21	0.371
Gender.Female	0.06	-0.35	0.46	0.769	0.13	-0.34	0.60	0.609	-0.09	-0.68	0.49	0.765
Age	0.13	0.04	0.23	0.008**	0.02	-0.09	0.13	0.717	-0.11	-0.24	0.02	0.121
Education	-0.05	-0.28	0.17	0.652	0.24	-0.02	0.50	0.093.	0.28	-0.05	0.61	0.112
Race.White	0.48	-0.05	0.96	0.067.	0.30	-0.26	0.87	0.320	-0.49	-1.18	0.20	0.191
Income	0.03	-0.02	0.07	0.234	-0.02	-0.07	0.03	0.493	-0.03	-0.09	0.04	0.418

Table 11. Linear effects models: the effect of language intensity and fact-checking agency on message intensity, message credibility, and engagement intentions (Mturk).

	Message intensity				Message credibility				Engagement			
	b	95% CI		p	b	95% CI		p	b	95% CI		p
		Lower	Upper			Lower	Upper			Lower	Upper	
Intercept	3.49	2.89	4.10	0.000	4.36	3.74	4.98	0.000	1.83	1.13	2.53	0.000
Intensity.Mid	0.01	-0.05	0.07	0.796	-0.06	-0.12	0.00	0.047*	0.01	-0.05	0.07	0.784
Intensity.High	0.00	-0.06	0.06	0.962	-0.01	-0.07	0.05	0.815	0.01	-0.06	0.07	0.871
Agency.Human	0.02	-0.11	0.15	0.765	0.06	-0.07	0.19	0.375	-0.04	-0.19	0.11	0.615
Topic.Vaccine	-0.03	-0.09	0.03	0.358	0.00	-0.06	0.06	0.925	0.03	-0.03	0.10	0.295
Topic.Election	-0.07	-0.12	-0.01	0.025*	-0.08	-0.14	-0.02	0.006**	-0.04	-0.10	0.02	0.221
AI familiarity	0.05	0.03	0.08	0.000***	0.05	0.02	0.08	0.001**	-0.04	-0.07	-0.01	0.015*
Conspiracy ideation	0.21	0.12	0.31	0.000***	0.08	-0.01	0.18	0.079.	0.57	0.46	0.68	0.000***
Party affiliation	-0.03	-0.06	0.00	0.036*	-0.04	-0.07	-0.01	0.006**	-0.07	-0.10	-0.04	0.000***
Political ideology	0.05	0.03	0.08	0.000***	0.04	0.01	0.06	0.008**	0.04	0.01	0.07	0.023*
Gender.Female	0.26	0.12	0.40	0.000***	0.19	0.05	0.33	0.008**	0.08	-0.08	0.24	0.356
Age	0.01	0.00	0.01	0.056.	0.00	0.00	0.01	0.348	0.00	-0.01	0.01	0.993
Education	0.04	-0.03	0.12	0.272	0.03	-0.05	0.11	0.474	0.21	0.12	0.30	0.000***
Race.White	-0.25	-0.40	-0.11	0.001**	-0.28	-0.43	-0.14	0.000***	-0.14	-0.31	0.02	0.090
Income	0.02	-0.01	0.04	0.247	0.01	-0.01	0.04	0.327	0.04	0.01	0.07	0.015*

Table 12. Linear effects models: the effect of language intensity and fact-checking agency on message intensity, message credibility, and engagement intentions (interaction added; Mturk).

	Message intensity				Message credibility				Engagement			
	b	95% CI Lower	95% CI Upper	p	b	95% CI Lower	95% CI Upper	p	b	95% CI Lower	95% CI Upper	p
Intercept	3.49	2.88	4.09	0.000	4.35	3.73	4.96	0.000	1.82	1.12	2.52	0.000
Intensity.Mid	-0.01	-0.09	0.08	0.873	-0.06	-0.15	0.02	0.157	0.02	-0.07	0.11	0.691
Intensity.High	0.03	-0.05	0.11	0.456	0.04	-0.04	0.13	0.340	0.02	-0.07	0.11	0.665
Agency.Human	0.03	-0.11	0.18	0.669	0.09	-0.06	0.24	0.231	-0.02	-0.19	0.14	0.789
Intensity.Mid:Human	-0.03	-0.09	0.03	0.353	0.00	-0.06	0.06	0.917	0.03	-0.03	0.10	0.295
Intensity.High:Human	-0.07	-0.13	-0.01	0.024*	-0.08	-0.14	-0.02	0.006**	-0.04	-0.10	0.02	0.219
Topic.Vaccine	0.05	0.03	0.08	0.000***	0.05	0.02	0.08	0.001**	-0.04	-0.07	-0.01	0.015*
Topic.Election	0.21	0.12	0.31	0.000***	0.08	-0.01	0.18	0.079.	0.57	0.46	0.68	0.000***
AI familiarity	-0.03	-0.06	0.00	0.036*	-0.04	-0.07	-0.01	0.006**	-0.07	-0.10	-0.04	0.000***
Conspiracy ideation	0.05	0.03	0.08	0.000***	0.04	0.01	0.06	0.008**	0.04	0.01	0.07	0.023*
Party affiliation	0.26	0.12	0.40	0.000***	0.19	0.05	0.33	0.008**	0.08	-0.08	0.24	0.356
Political ideology	0.01	0.00	0.01	0.056.	0.00	0.00	0.01	0.348	0.00	-0.01	0.01	0.993
Gender.Female	0.04	-0.03	0.12	0.272	0.03	-0.05	0.11	0.474	0.21	0.12	0.30	0.000***
Age	-0.25	-0.40	-0.11	0.001**	-0.28	-0.43	-0.14	0.000***	-0.14	-0.31	0.02	0.090.
Education	0.02	-0.01	0.04	0.247	0.01	-0.01	0.04	0.327	0.04	0.01	0.07	0.015
Race.White	0.03	-0.09	0.15	0.631	0.00	-0.12	0.12	0.980	-0.02	-0.15	0.11	0.772
Income	-0.07	-0.18	0.05	0.274	-0.10	-0.21	0.02	0.116	-0.03	-0.16	0.10	0.654

Table 13. Linear regression models: the effect of fact-checking agency on agency evaluation (Mturk).

	b	95% CI		p
		Lower	Upper	
Intercept	4.61	3.90	5.31	0.000
Agency.Human	0.02	-0.14	0.17	0.819
AI knowledge	0.05	0.02	0.08	0.003**
Conspiracy ideation	0.10	-0.01	0.21	0.077.
Party affiliation	-0.05	-0.09	-0.01	0.009**
Political ideology	0.03	0.00	0.06	0.084.
Gender.Female	-0.18	-0.34	-0.02	0.031*
Age	0.00	-0.01	0.01	0.495
Education	0.02	-0.07	0.11	0.656
Race.White	-0.22	-0.39	-0.05	0.010*
Income	0.03	0.00	0.00	0.041*

Table 14. Linear effects models: the effect of language intensity and fact-checking agency on message intensity, message credibility, and engagement intentions (counter-attitudinal posts; Mturk).

	Message intensity				Message credibility				Engagement			
	b	95% CI		p	b	95% CI		p	b	95% CI		p
		Lower	Upper			Lower	Upper			Lower	Upper	
Intercept	2.60	1.71	3.49	0.000	3.14	2.20	4.05	0.000	0.60	-0.35	1.55	0.220
Intensity.Mid	0.04	-0.13	0.22	0.628	0.02	-0.16	0.21	0.810	0.01	-0.18	0.20	0.918
Intensity.High	0.10	-0.08	0.27	0.285	0.09	-0.10	0.27	0.348	-0.04	-0.23	0.16	0.721
Agency.Human	0.06	-0.12	0.24	0.538	0.14	-0.05	0.32	0.156	-0.03	-0.23	0.16	0.726
Topic.Vaccine	-0.10	-0.31	0.12	0.359	0.02	-0.20	0.24	0.889	0.17	-0.06	0.41	0.158
Topic.Election	-0.04	-0.25	0.17	0.689	-0.07	-0.29	0.16	0.533	0.24	-0.01	0.49	0.046*
AI familiarity	0.03	-0.01	0.07	0.222	-0.02	-0.06	0.02	0.303	-0.08	-0.12	-0.04	0.000***
Conspiracy ideation	0.32	0.17	0.46	0.000***	0.23	0.08	0.38	0.003**	0.71	0.55	0.86	0.000***
Party affiliation	-0.01	-0.05	0.03	0.663	-0.03	-0.07	0.01	0.149	-0.07	-0.11	-0.03	0.002**
Political ideology	0.10	0.06	0.14	0.000***	0.08	0.04	0.12	0.000***	0.06	0.02	0.10	0.006**
Gender.Female	0.23	0.03	0.43	0.027*	0.16	-0.05	0.37	0.135	0.02	-0.20	0.23	0.878
Age	0.01	0.00	0.01	0.265	0.00	-0.01	0.01	0.938	0.00	-0.01	0.01	0.928
Education	0.07	-0.04	0.18	0.233	0.11	0.00	0.23	0.053.	0.30	0.19	0.42	0.000***
Race.White	-0.19	-0.39	0.02	0.079.	-0.17	-0.38	0.05	0.126	-0.12	-0.34	0.10	0.293
Income	0.00	-0.03	0.04	0.890	-0.01	-0.05	0.03	0.704	0.02	-0.02	0.06	0.248

Table 15. Linear effects models: the effect of language intensity and fact-checking agency on message intensity, message credibility, and engagement intentions (pro-attitudinal posts; Mturk).

	Message intensity				Message credibility				Engagement			
	b	95% CI Lower	95% CI Upper	p	b	95% CI Lower	95% CI Upper	p	b	95% CI Lower	95% CI Upper	p
Intercept	3.63	2.99	4.26	0.000	4.49	3.85	5.13	0.000	2.16	1.42	2.91	0.000
Intensity.Mid	0.01	-0.07	0.09	0.786	-0.05	-0.14	0.03	0.185	0.00	-0.09	0.08	0.931
Intensity.High	-0.03	-0.11	0.06	0.534	-0.04	-0.12	0.04	0.336	-0.04	-0.13	0.05	0.416
Agency.Human	0.03	-0.11	0.17	0.653	0.06	-0.08	0.20	0.383	-0.04	-0.20	0.12	0.668
Topic.Vaccine	0.01	-0.07	0.08	0.859	0.00	-0.07	0.07	0.954	0.02	-0.05	0.10	0.544
Topic.Election	-0.09	-0.18	0.00	0.059.	-0.06	-0.16	0.03	0.172	-0.14	-0.24	-0.04	0.007**
AI familiarity	0.06	0.03	0.09	0.000***	0.06	0.03	0.09	0.000***	-0.03	-0.07	0.00	0.050.
Conspiracy ideation	0.19	0.09	0.29	0.000***	0.05	-0.05	0.14	0.359	0.50	0.39	0.62	0.000***
Party affiliation	-0.03	-0.07	0.00	0.027*	-0.03	-0.06	0.00	0.051	-0.06	-0.09	-0.02	0.002**
Political ideology	0.05	0.02	0.08	0.000***	0.04	0.01	0.07	0.005***	0.05	0.02	0.08	0.004**
Gender.Female	0.28	0.13	0.42	0.000***	0.23	0.08	0.37	0.003***	0.07	-0.10	0.24	0.404
Age	0.01	0.00	0.01	0.102	0.00	0.00	0.01	0.187	0.00	-0.01	0.01	0.909
Education	0.05	-0.03	0.13	0.202	0.02	-0.06	0.10	0.677	0.19	0.09	0.28	0.000***
Race.White	-0.25	-0.40	-0.10	0.001**	-0.32	-0.47	-0.17	0.000**	-0.15	-0.33	0.02	0.092.
Income	0.01	-0.02	0.04	0.422	0.01	-0.02	0.04	0.481	0.04	0.00	0.07	0.026*

Table 16. Linear effects models: the effect of language intensity and fact-checking agency on message intensity, message credibility, and engagement intentions (no-attitude posts; Mturk).

	Message intensity				Message credibility				Engagement			
	b	95% CI		p	b	95% CI		p	b	95% CI		p
		Lower	Upper			Lower	Upper			Lower	Upper	
Intercept	3.44	2.25	4.63	0.000	3.73	2.56	4.92	0.000	1.30	0.03	2.56	0.050
Intensity.Mid	0.04	-0.13	0.21	0.639	-0.15	-0.34	0.05	0.152	0.08	-0.19	0.32	0.570
Intensity.High	0.09	-0.09	0.27	0.330	0.13	-0.07	0.33	0.214	0.13	-0.13	0.39	0.326
Agency.Human	-0.04	-0.29	0.22	0.784	0.00	-0.25	0.25	0.993	0.13	-0.14	0.39	0.353
Topic.Vaccine	0.08	-0.10	0.27	0.385	0.22	0.01	0.43	0.040*	0.08	-0.18	0.36	0.549
Topic.Election	-0.05	-0.23	0.13	0.573	-0.10	-0.31	0.10	0.322	0.22	-0.04	0.49	0.115
AI familiarity	0.08	0.02	0.14	0.010*	0.08	0.02	0.14	0.008**	0.00	-0.06	0.06	0.885
Conspiracy ideation	0.16	-0.07	0.38	0.181	0.18	-0.04	0.40	0.109	0.49	0.26	0.73	0.000***
Party affiliation	-0.05	-0.12	0.01	0.117	-0.05	-0.12	0.01	0.103	-0.07	-0.14	0.00	0.044*
Political ideology	0.08	0.02	0.13	0.007**	0.07	0.01	0.12	0.015*	0.02	-0.03	0.08	0.400
Gender.Female	0.12	-0.15	0.39	0.395	0.10	-0.16	0.37	0.451	0.26	-0.03	0.53	0.082.
Age	0.00	-0.01	0.02	0.519	0.00	-0.01	0.02	0.556	0.00	-0.01	0.01	0.902
Education	0.01	-0.13	0.15	0.871	-0.04	-0.18	0.09	0.555	0.23	0.09	0.38	0.003**
Race.White	-0.12	-0.42	0.17	0.428	0.00	-0.29	0.30	0.979	0.10	-0.22	0.41	0.554
Income	0.00	-0.05	0.06	0.898	-0.01	-0.07	0.04	0.608	0.00	-0.05	0.06	0.933

Figure 1. Mean differences of message intensity, message credibility, and engagement intentions across language intensity and fact-checking agency (SONA).

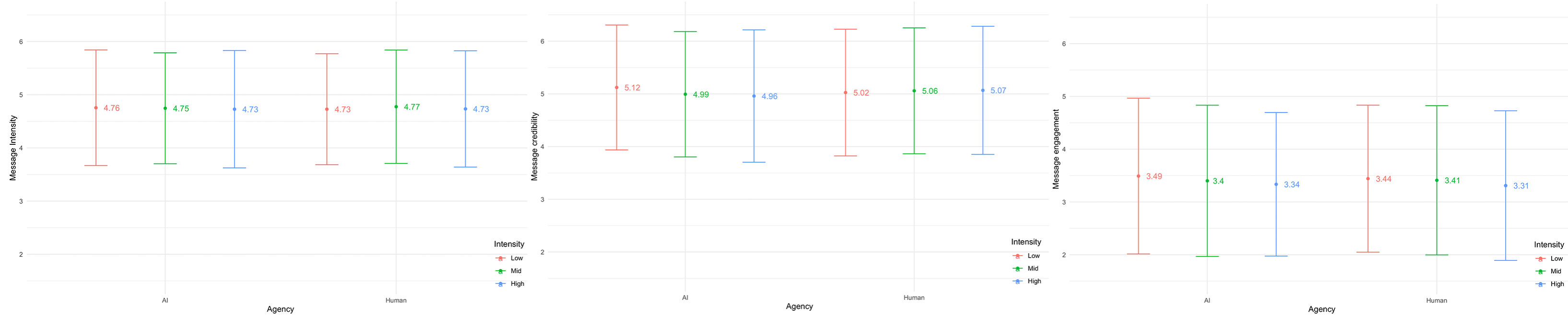

Figure 2. Mean differences of message intensity, message credibility, and engagement intentions across language intensity and fact-checking agency (Mturk).

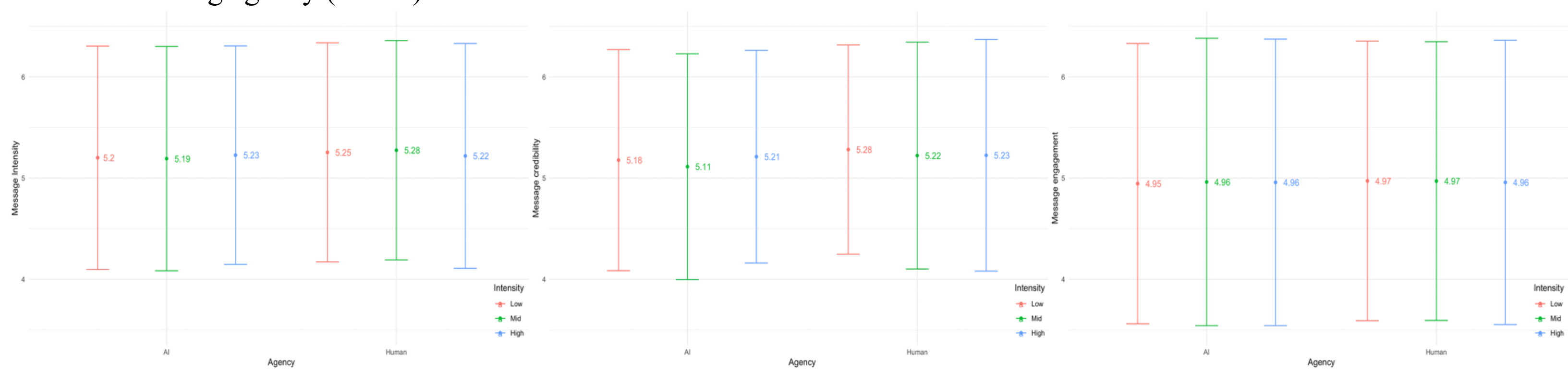

Figure 3. Mean differences of message intensity, message credibility, and engagement intentions across topics (SONA).

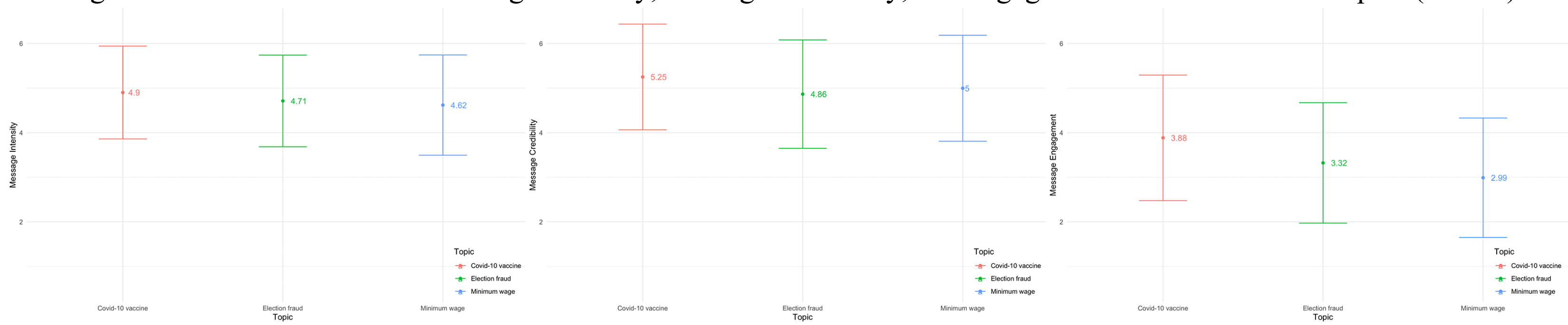

Figure 4. Mean differences of message intensity, message credibility, and engagement intentions across topics (Mturk).

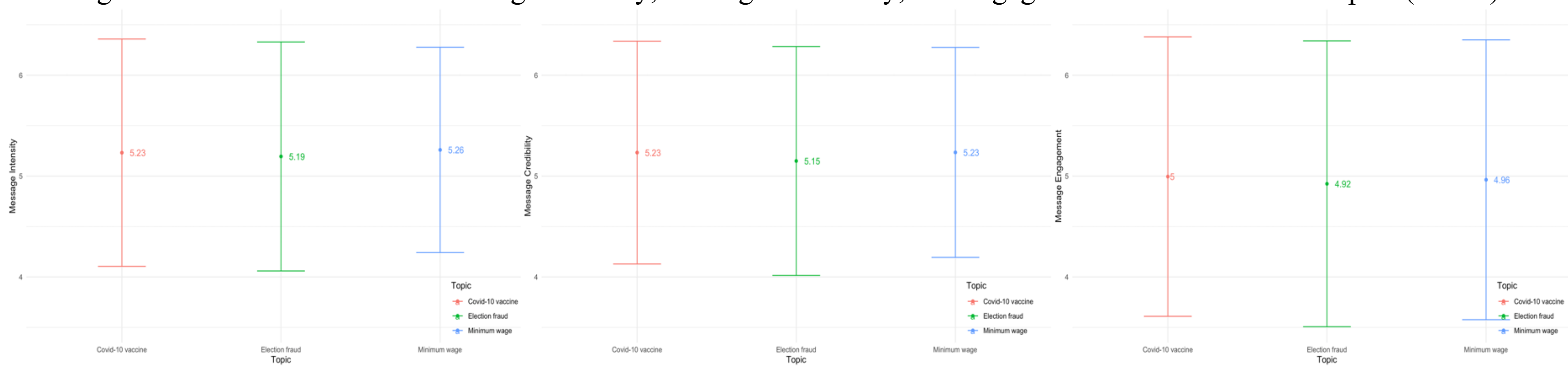

Figure 5. Sample fact-checking messages in AI and human conditions.