Images, Emotions, and Credibility: Effect of Emotional Facial Expressions on Perceptions of News Content Bias and Source Credibility in Social Media

Alireza Karduni[1], Ryan Wesslen[2], Douglas Markant[2], Wenwen Dou[2]

[1]IDEO
[2]University of North Carolina at Charlotte
alireza116@gmail.com, rwesslen@gmail.com, dmarkant@uncc.edu, wdou1@uncc.edu

Abstract

Images are an indispensable part of the news we consume. Highly emotional images from mainstream and misinformation sources can greatly influence our trust in the news. We present two studies on the effects of emotional facial images on users' perception of bias in news content and the credibility of sources. In study 1, we investigate the impact of repeated exposure to content with images containing positive or negative facial expressions on users' judgments of source credibility and bias. In study 2, we focus on sources' systematic emotional portrayal of specific politicians. Our results show that the presence of negative (angry) facial emotions can lead to perceptions of higher bias in content. We also find that consistent negative portrayal of different politicians leads to lower perceptions of source credibility. These results highlight how implicit visual propositions manifested by emotions in facial expressions may substantially affect our trust in news.

Introduction

In the aftermath of the US 2020 presidential election, some rioters who did not trust the official results of the election stormed the capitol building. Around the same time, when the first Covid-19 vaccines were introduced, many did not trust the news about the safety of the vaccines. Observing such incidents, what causes news consumers not to trust certain mainstream news sources while believing misinformation published by others? Recent research suggests one answer may be cognitive factors, such as the propensity for analytical reasoning affecting belief in the accuracy of news content (Pennycook and Rand 2021). However, our judgments of news source credibility is also an important but often neglected factor shaping our judgments of news and misinformation (Pehlivanoglu et al. 2021; Wallace, Wegener, and Petty 2020b). Indeed, prominent misinformation scholars advocate for increasing focus on behaviors of news sources rather than individual stories since "the defining element of fake news is the intents and processes of a publisher" (Lazer et al. 2018). In this research, we tackle the interactions between news sources and consumers by exploring how social media users' evaluations of the news content influence their judgments of news source credibility. We posit

that identifying factors that influence consumers' judgments of news sources is essential, not only to help detect and combat misinformation but also to identify and combat practices that reduce consumers' trust in trustworthy sources.

News sources often frame their content with specific styles, tones, and emotions. Both mainstream news media and misinformation sources use highly emotionalized and sensational content to influence their audiences (Arsenault and Castells 2006; Wardle and Derakhshan 2017; Volkova et al. 2017; Richards 2007; Peng 2018; Rathje, Van Bavel, and van der Linden 2021). Various studies have shown that such content strategies are often effective. People are more likely to be drawn to highly emotionalized news, click on articles associated with extreme emotions, and share headlines with more negative sentiments (Bowman and Cohen 2020; Reis et al. 2015; Berger and Milkman 2012). In addition to emotional verbal content, news sources take advantage of social media's highly visual nature by choosing images that amplify the persuasive power of the verbal content (Wardle and Derakhshan 2017; Grabe and Bucy 2009). Examples of such visual content include images that communicate racist concepts not present within the text (Messaris and Abraham 2001), emotional facial expressions of news-casters (Mullen et al. 1986), and partisan media producing ideological bias in their visual coverage of political candidates with features including facial expressions and face size (Peng 2018).

In this work, we tackle a salient aspect of visual information, facial expressions, in social media posts on users' judgments: Do users perceive a source as less credible when the source repeatedly publishes tweets with angry facial expressions? What if a source systematically portrays a negative visual bias against a politician (e.g. Bernie Sanders or Donald Trump) by repeatedly publishing news with angry images of those politicians? We explore these questions by considering **source credibility**, a primary factor in the persuasiveness of news sources (Hovland, Janis, and Kelley 1953; Wallace, Wegener, and Petty 2020b); and **content bias**, defined as "having a perspective that is skewed," (Wallace, Wegener, and Petty 2020b) as two primary elements of users' judgments of news sources. We investigate how repeated usage of images with positive (happy) or negative (angry) emotional facial expressions in social media posts affects users' judgments about **source credibility** and **content bias**.

We seek to answer the following questions: 1) How does repeated exposure to news content accompanied by images with positive (happy) or negative (angry) facial expressions impact users' judgments about a source's credibility?; and 2) How sources' repeated positive or negative visual portrayal of different politicians interacts with users' prior attitudes (favorability and familiarity) towards those politicians to shape their judgments of source credibility. To answer these questions, we conducted two consecutive preregistered controlled experiments:

1. In study 1, we examined how users' credibility judgments are impacted by news sources that show a pattern of using highly emotionalized visual images.[1] More specifically, we investigated how angry or happy facial expressions in tweet images affect users' judgments about eight anonymized right/left-leaning and mainstream/misinformation news accounts.

2. Leveraging findings from study 1, in study 2, we focused on sources with repeated emotionalized visual coverage of influential politicians.[2] We examined how users' judgments are impacted by sources that consistently portray specific politicians' emotions as negative or positive. In study 2, we also investigated how users' prior attitudes towards each politician interact with sources' negative or positive emotional portrayals of those politicians.

In study 1, we observe that users find tweets with negative (angry) facial expressions as more biased. However, we found that negative facial expressions might not lead to lower perceived source credibility and that this effect is likely heavily moderated by textual content and tone of the news coverage (e.g., misinformation sources were deemed as less credible). In study 2, we found that focusing on mainstream sources' repeated negative visual portrayal of specific politicians, users are more likely to perceive the content as more biased and the source as less credible. Across both studies, in addition to quantitative analysis of the experimental treatments, we qualitatively analyzed users' comments explaining the reasons behind their judgments. Our qualitative analysis demonstrated that users also relied on several cues such as opinionated language, political bias, and satire, to guide their credibility judgments. These studies highlight the extent to which highly emotionalized visual coverage might impact users' trust in news sources.

Related Work

Source Credibility and Content Bias. False, inaccurate or "fake" news is a serious problem for our societies and democracies (Lazer et al. 2018; Karduni et al. 2019; Wardle and Derakhshan 2017). In their work on decision-making about misinformation, Pennycook and Rand highlight different aspects of why users fall for "fake news." They provide evidence that "contradict the common narrative that partisanship and politically motivated reasoning explain why people fall for fake news" ((Pennycook and Rand 2021) pg. 1) Furthermore, they find that the propensity to engage in

reasoning is a more important factor in users' ability to detect false news.

In addition to users' ability to detect instances of inaccurate and false information, perceptions of source credibility influence how users interact with and are persuaded by the news (Pornpitakpan 2004). Perceptions of trustworthiness and expertise are noted as two main factors in credibility judgments (Petty, Wegener, and Fabrigar 1997; Wallace, Wegener, and Petty 2020b). Research on credibility perceptions on social media has highlighted that when faced with short content (e.g., Twitter), users might have different credibility perceptions for different topics (e.g. science versus entertainment) and might rely on different heuristics such as username style to judge source credibility (Morris et al. 2012). Recent work by Laura Wallace suggests that perceptions of source bias and untrustworthiness can have different effects on perceptions of source credibility. For instance, in certain cases, users might believe sources are biased but honest in their reporting and therefore credible (Wallace, Wegener, and Petty 2020b,a).

Images, Emotions, and Users' Judgments of News. Visual information impacts how users judge the credibility of information sources in several ways. Wobbrock et al. show that visual appearances like font size or inclusion of images may impact users' perceptions of credibility (Wobbrock et al. 2021). Different images may have varying effects on users' judgments. For example, after seeing an image of a brain, users are more likely to believe claims in certain scientific articles (Schweitzer, Baker, and Risko 2013; McCabe and Castel 2008) while images of smoking increases belief in messages in warning signals (Shi et al. 2017). Articles accompanied by alarming images (Knobloch et al. 2003) or ones that depict victimization (Zillmann, Knobloch, and Yu 2001) increase users' selective interaction with the articles. Furthermore, when exposed to false versus true news headlines, participants' self-reported experience of heightened emotion (users who report feeling more angry or sad) is linked to an increase in perceived accuracy of false news but not factually correct news (Martel, Pennycook, and Rand 2020).

Recent work in social psychology and media bias has highlighted the interpersonal effect of emotions on users' decisions (Van Kleef, De Dreu, and Manstead 2010). For example, displays of angry (but not happy) facial expressions might reduce the likelihood of our decisions to trust others (Campellone and Kring 2013). In the context of charity advertisements, images with sad facial expressions can increase the likelihood of donating (Small and Verrochi 2009). Moreover, displays of different emotions can impact perceptions of cooperativeness and appraisal (e.g., happy agent perceived to be more likely to cooperate and angry/sad the opposite) (de Melo et al. 2012). Andalibi and Buss assert that people find emotions as "insights to behavior, prone to manipulation, intimate, vulnerable, and complex" (Andalibi and Buss 2020). Emotional content in images also impacts the likelihood of people believing a statement. For example, exposure to highly negative emotional images about different phenomena increases belief in the accuracy of news con-

tent (Vlasceanu, Goebel, and Coman 2020). In comparison to positive imagery, being exposed to negative imagery has been associated with greater perceived accuracy of false information (Porter et al. 2010). Melodramatic animations are also shown to likely increase users' perceptions of the credibility of a news report (Lo and Cheng 2017). From the perspective of visual media bias, (Grabe and Bucy 2009) called for more attention on studying the visual framing of politicians in the news due to the increase in visual (image) portrayals in the news. A study by (Peng 2018) showed that, among other features, happy facial images increase the perceived favorability of a political candidate and perceived trustworthiness, but angry facial expressions decrease both. Another study by Masch and Gabriel suggests that depending on the political context, positive and negative depictions of politicians in videos might impact users' attitudes towards those politicians (Masch and Gabriel 2020).

Current research on misinformation focuses on perceptions of content accuracy and users' ability to distinguish false and true information (Pennycook and Rand 2021) based on various cues, including emotions in text (Karduni et al. 2018) or images (Vlasceanu, Goebel, and Coman 2020). Research on persuasion has shown users' judgments of source credibility and content bias, and the certainty of those judgments, are important factors in news sources' persuasive power (Wallace, Wegener, and Petty 2020b; Tormala and Petty 2004). Inspired by the mentioned work, this paper studies an important but, to our knowledge, overlooked aspect of our judgments of news sources: the effect of sources' repeated and potentially systematic usage of highly emotional facial images on users' judgments of source credibility and content bias.

Study Design and Implementation

Considering perception of source credibility as a product of the news content that users consume, we formulate the overall structure of the studies[3] as 1) Users first view curated social media posts from an **anonymized** news source; 2) Users evaluate bias for multiple posts from the source; 3) After deciding they have viewed and evaluated enough posts from a source (with a hard minimum of five tweets), users are instructed to provide a credibility rating of the source. Users repeated these steps for eight different new sources.

This structure applies to both studies 1 and 2. The randomized treatments were applied by altering the content users see from each source (i.e., content with positive or negative images or no images as a control condition).

As opposed to limiting users' judgments to binary outcomes, we elicited users' bias and credibility judgments on continuous response scales. We elicited perceived bias of a tweet as a continuous number between 0 (unbiased) and 1 (biased). Similarly, we elicited perceived credibility of a source as a number between 0 (not credible) and 1 (credible). As it has been argued that user attitudes with higher certainties have important consequences, including increasing resistance to persuasion and being persistent over time (Tormala and Petty 2004; Tormala and Rucker 2007), we

---

[3]Approved by our institutions' Internal Review Board (IRB)

elicited users' certainty around their judgments as a confidence interval within each respective domain. We adopted a modified version of the Line + Cone technique (Karduni et al. 2020) to elicit users' judgments. The visual technique (called Line + Range; See Figure 7) enabled us to elicit judgments and uncertainty using a visual technique. To explore users' self-described reasons behind their judgments, we also asked users to provide verbal comments using an open text box.

Dataset and Study Stimuli

The tweets dataset was collected from the Twitter streaming API from October 25th to January 25th, 2018, from multiple news accounts labeled as misinformation or mainstream by Karduni et al. (Karduni et al. 2019, 2018). After downloading all images from the dataset, we used the python face-recognition library[4] to extract images with faces and cropped the images only to include faces. To identify images of different politicians, we used Google's FECNet (Schroff, Kalenichenko, and Philbin 2015) to extract feature vectors of cropped faces and used HDBScan (McInnes, Healy, and Astels 2017) to cluster faces and manually identified influential politicians. We chose to include only cropped faces in our study to remove potential confounding factors such as body language and surrounding scenes, as studies show that recognition of facial expressions is influenced by both(Kret et al. 2013; Kret and Gelder 2012, 2013).

To extract emotion predictions from faces, we used two separate python libraries, EmoPy [5] and deepface[6], which provide the ability to sort the tweets based on emotion predictions and model confidence. After generating a candidate dataset, we qualitatively excluded low-resolution images, inaccurately assigned to a cluster, or had wrong emotion predictions, All qualitative coding to produce the final datasets for both studies was done by the authors. The first author first created a larger candidate dataset. Using the study interface, the other authors iteratively evaluated each data point. Disagreements were resolved in team meetings.

Study 1 & 2 Models

Our experimental design for both studies consisted of repeated measures for each user within two levels of responses, including source-level responses (credibility) and tweet-level responses (bias). Users' responses were continuous variables bounded between 0 and 1. For both studies, we used mixed-effects beta regressions with `glmmTMB` to address the hierarchical design of our studies and the bounded dependent variables. Within the text, model coefficients are reported as log-odds ratios with corresponding confidence intervals. In the model figures, we transformed the log-odds ratios to odds ratios for ease of interpretation. Odds ratios show the direction and strength of how each independent variable impacts the dependent variables. We used the normal approximation to calculate p-values of fixed effects and t-values produced by `lme4`.
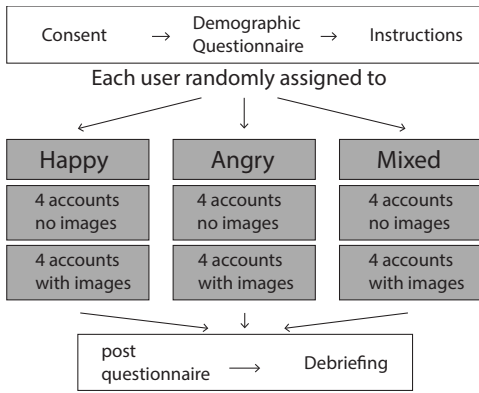
---

[4]https://pypi.org/project/face-recognition/
[5]https://github.com/thoughtworksarts/EmoPy
[6]https://pypi.org/project/deepface/

Figure 1: Study 1 conditions and process

Source name	Type	Orientation
@veteranstoday	misinformation	left
@amlookout	misinformation	right
@opednews	misinformation	left
@InvestWatchBlog	misinformation	right
@MotherJones	mainstream	left
@Jeresulem_Post	mainstream	right
@cnnPolitics	mainstream	left
@nypost	mainstream	right

Table 1: 8 anonymized Twitter accounts in Study 1

Study 1

In Study 1, we examined how positive (happy) and negative (angry) emotions in images influence perceived content bias (tweets) and source credibility (accounts). Participants completed a task in which they observed a series of tweets from eight different accounts. To provide a variety of different political orientations and content types, the accounts were either categorized as **mainstream** or known to produce **misinformation**. The accounts were also categorized as being right-leaning or left-leaning (see Table 1).[7] Given these two dimensions of mainstream/misinformation and right/left political orientation, users evaluate a total of eight accounts with more than 90 tweets in the final dataset. The content shown to users was manipulated in a between-subjects manner. Users were randomly assigned to one of three conditions: **happy**, **angry**, and **mixed** (see Figure 1). The happy condition contained tweets with the highest happiness scores for the images. The angry condition included tweets containing images with the highest angry image scores. The mixed (control) condition alternates between happy and angry images. Furthermore, one account from each category (right, left, misinformation, mainstream) was randomly selected to include no images. For example, a participant assigned to a happy condition, viewed tweets with the highest-rated happy images from four of the eight accounts but did not see images from the other four accounts.

For each account, users responded based on their judgments about the bias of each tweet. By clicking on "View more tweets", participants viewed additional tweets until they decided to rate the source's credibility (a minimum of 5 tweets was enforced for each account). By clicking on "Make a decision," users saw a pop-up view in which they entered their credibility judgments. Users also had the option to use a text box to describe what influenced their decisions.

Hypotheses

We hypothesize that in comparison to tweets with happy images, users are more likely to assess tweets with angry images as biased. Furthermore, we also hypothesize that as

compared to the mixed condition, users in the angry condition are more likely to assess sources as less credible. However, since it is likely for users' judgment to be influenced by the inherent differences in the nature of sources, we also explore the effects of political orientation (right vs. left) and source type (mainstream vs. misinformation) on their judgments.

Dependent & Independent Variables

We considered four total dependent variables (DV): (1) content bias (bounded value between [0,1]), (2) uncertainty range around content bias (bounded value between [0,1]), (3) source credibility (bounded value between [0,1]), (4) uncertainty around source credibility (bounded value between [0,1]). For our independent variables (IV), we included the image emotion condition (angry, happy, or mixed), image shown (true or false), as well as political orientation (right or left), and source type (mainstream or misinformation). For models built based on tweet-level responses, i.e. bias choice/uncertainty of tweets as the dependent variable, only two image emotion conditions of happy or angry are applicable since each tweet contains at most one type of emotion.

Model Specification

For each model, we included users' unique IDs and the source id as random effects. The reference conditions for credibility choice and uncertainty models are image emotion = mixed, source orientation = left, source type = mainstream, and the image shown = False. The reference conditions for bias choice and uncertainty models are image emotion = happy, source orientation = left, source type = mainstream, and the image shown = False.
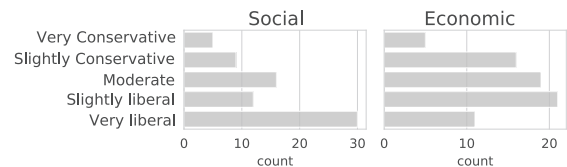

Figure 2: Participants' responses to questions "How would you describe your political outlook with regard to economic/social issues?" for Study 1.

---

[7]Right or left-leaning labels were confirmed with https://mediabiasfactcheck.com/ and https://www.allsides.com/

Bias Choice

(Intercept) 0.73 ***
image_emotion [angry] 1.25 ***
source_orientation [right] 1.24 **
source_type [misinformation] 1.61 ***
image_shown [with image] 1.13

Odds Ratios: 0.5 1.0 1.5 2.0

Bias Uncertainty

(Intercept) 0.37 ***
1.04
1.14 **
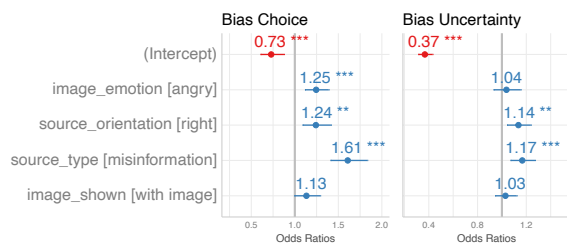1.17 ***
1.03

Odds Ratios: 0.4 0.8 1.2

Figure 3: Study 1 fixed effects odds ratios for bias choice (left) and bias uncertainty (right). Error bars indicate 95% confidence intervals. Asterisks indicate statistical significance than zero using p-values: *** 99.9%, ** 99%, * 95%.

Credibility Choice

(Intercept) 1.30 *
image_emotion [happy] 0.87
image_emotion [angry] 0.94
source_orientation [right] 0.77 **
source_type [misinformation] 0.58 ***
image_shown [with image] 0.87

Odds Ratios: 0.5 1.0 1.5 2.0

Credibility Uncertainty

(Intercept) 0.31 ***
1.21
1.07
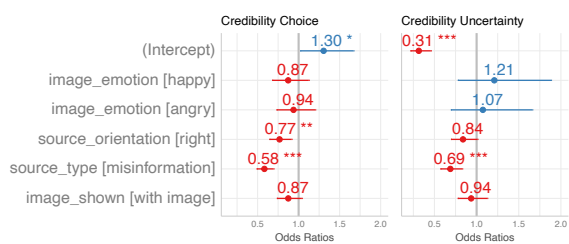0.84
0.69 ***
0.94

Odds Ratios: 0.5 1.0 1.5 2.0

Figure 4: Study 1 fixed effects coefficients for credibility choice (left) and credibility uncertainty (right). Error bars indicate 95% confidence intervals. Asterisks indicate statistical significance than zero using p-values: *** 99.9%, ** 99%, * 95%.

Participants

In this study's pre-registration, we originally planned to recruit up to 300 participants from crowdsourcing platforms. However, due to concerns about the quality of data collected from MTurk and time constraints, we shifted our participant pool to university students. We recruited a total of 81 university students with an average age of 21 years old (std 5.6). Per our pre-registration, we excluded responses from 9 participants with missing responses (due to unexpected technical difficulties), resulting in 72 accepted responses (48 women, 23 men, and 1 other; 40 white, 13 African American, 7 other Asian, 4 East Asian, 5 Hispanic, 2 middle eastern, and 1 native American). Participants were mostly liberal in regards to social, political issues and more balanced in regards to economic issues (see Fig 2). Per random assignment, 30 participants were assigned to the angry condition, 24 participants to the happy condition, and the remaining 18 to the mixed condition. Participants took an average of 26 minutes to complete the study. All participants either received course extra credits or research credits.

Results

Content Bias. We used two mixed-effects beta regressions to study the effects on users' judgments about the bias of individual tweets (see Fig 3-left). Users found content from right-leaning sources to be more biased ($\beta = 0.216$ $[0.086, 0.34]$, $z = 3.25, p < .01$). Similarly, users judged content from misinformation sources to be more biased ($\beta = 0.474$ $[0.34, 0.60]$, $z = 7.11, p < .001$). We also found that users rated tweets containing angry images as more biased ($\beta = 0.219$ $[0.11, 0.32]$, $z = 3.94, p < .001$). This is in line with our hypothesis that angry imagery will lead to an increase in users' perceived bias. Surprisingly, even though the content shown to users was solely sorted based on facial emotions in images, we did not find a noticeable effect of emotional images on users' bias ratings. It is possible that sources choose images to amplify tweets' content, and users' judgments were not based on the emotion in facial expressions alone.

In our mixed-effects model on users' uncertainty around their judgments (see Fig 3-right), we found that compared to left-leaning mainstream accounts, users were more likely to have larger uncertainty ranges for right-leaning accounts

($\beta = 0.127$ $[0.041, 0.21]$, $z = 2.90$ $p < .01$). We also found a similar positive effect for misinformation accounts ($\beta = 0.154$ $[0.06, 0.24]$, $z = 3.52, p < .001$).

Source Credibility. We used mixed-effects beta regression to investigate the effects of study conditions on users' judgments about source credibility (see Fig 4-left). We found that in reference to left-leaning accounts, users were more likely to rate right-leaning accounts as less credible ($\beta = -0.264$ $[-0.43, -0.08]$, $z = -2.961, p < .01$). Users were also more likely to rate misinformation sources as less credible ($\beta = -0.539$ $[-0.71, -0.36]$, $z = -6.024, p < .001$). We did not observe significant effects for the happy or angry conditions. In other words, we did not find evidence towards our hypotheses that being exposed to a source that repeatedly publishes tweets with angry or happy facial expressions impacts perceptions of source credibility.

We also used a mixed-effects beta regression to study whether users' confidence around their decisions (represented through uncertainty ranges in the Line + Range technique). Interestingly, source type was the only factor that significantly affected users' uncertainty about source credibility. Users were likely to be more confident in their decisions when source type was categorized as misinformation ($\beta = -0.373$ $[-0.55, -0.18]$, $z = -3.921, p < .001$).

Qualitative Analysis of Users' Comments

Each user had the option to answer one open-ended question for each account, asking "please describe how the tweets (text and images) influenced your decisions about this account?" Even though leaving comments was an optional part of the study, we received comments from all 72 participants, and the majority of participants left comments for all 8 trials. In total, we collected 572 comments about users' decision-making influences. Since thematic analysis with a large number of comments was challenging, we used topic modeling to facilitate the qualitative analysis of the comments and arrive at different themes of influences on users' judgments. We categorized the extracted topics into four general themes: 1) sources' political orientation (left-leaning or right-leaning), 2) opinionated versus factual reporting, 3) specific language usage or tone (e.g. source is angry or uses tabloid-like language), and 4) the effect of images. Although
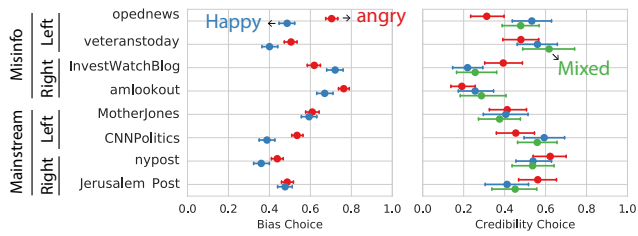
Figure 5: Mean and 95% bootstrapped confidence interval for bias and credibility choice for different news accounts (sources) in study 1.

the majority of users' comments were related to different cues tied to the text of the tweets, in this section, we provide a summary of comments related to images.

Topics 9 (25 comments from 14 users) and 5 (19 comments from 10 users) included comments that mentioned images as a factor in their decision. These comments made a wide range of image-related observations, from unflattering imagery to facial expressions and lack of seriousness. For example, a user described how they perceived the image and the text to be not aligned: *"Some of the tweets seemed to be about data rather than opinions. Some of the pictures seemed to take away merit."* Another comment explicitly mentioned facial expressions of individuals helping them decide that the tweets have a left bias: *"Many of the tweets seemed to be credible as most were quotes by others. Some of the pictures had facial expressions that made the tweets seem left swinging."*

A number of comments cited comic or not serious imagery as the basis for their decisions. For example, *"Difficult to take the meme-like images seriously."* And, *"The images were cartoonish and difficult to take it seriously. It was obviously making fun of Trump.".* A few comments mentioned unprofessional or unflattering images as the basis for their judgments: *"[...] and somewhat unflattering pictures of those who either commented against them or who may do (or not do) something against them."* And, *"Some of the pictures were not professional and showed the president and alliances in a negative light."*

These comments show that, at least for a group of users, images can serve as a cue for a source's lack of credibility. However, the majority of comments about pictures and images did not specifically mention emotions in images as the basis for their decisions. This might be due to the fact that tweets in our dataset contained a diverse set of topics. It is possible that users look for a more systematic negative treatment of specific topics or individuals rather than a combination of negative imagery from a wide range of topics.

Study 1 Discussion

We hypothesized that sources' systematic posting of tweets with angry images would lead to higher perceived content bias and lower perceived source credibility. We also hypothesized that happy images would lead to a reverse effect. We found partial evidence for our hypotheses: we observed an increase in perceived bias for tweets containing
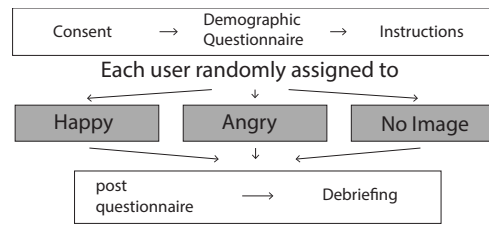

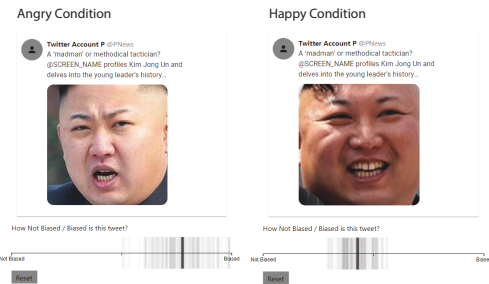Figure 6: Study 2 conditions and process

Figure 7: Sample of study 2 Angry (left) and Happy (right) stimuli for one tweet related to Kim Jong Un.

angry images (see Fig 5 for raw point estimates for each account). We did not find evidence that this perceived content bias led to lower judgments of source credibility. We believe this is in line with recent findings by Wallace and colleagues asserting that higher perceived bias does not necessarily lead to lower perceived credibility of sources (Wallace, Wegener, and Petty 2020b). However, perceptions of content bias were likely affected by the general tone and topical focus of individual accounts (note that $source\_type = misinformation$ had the largest effect on perceived bias. Our qualitative analysis of comments also hints that users rely heavily on text content to judge both bias in content and the credibility of sources. In retrospect, the topical diversity in the first study, both between different sources and within tweets published by a single source, does not necessarily portray systematic biases which would reduce the perceived credibility of a source. In the next study, we explore this issue by fixing the textual content shown to users and manipulating only the emotional valence of facial expressions in images.

Study 2

In study 1, the content shown to users was not limited to specific topics and we found that users' judgments were likely influenced by different topical focuses of sources. To better evaluate the impact of emotional facial expressions on users' judgments, we shifted the focus of study 2 to explore the effects of mainstream sources' systematic emotional portrayal of specific politicians. To address such a scenario, in all experimental conditions of this study 2, users received the same 8 sets of tweet texts about 8 different politicians. Per random assignment, users received either angry or happy images of the politicians. We included no images as a control condition in which users received the same sets of tweets without any accompanying images of those politicians.

Politician	Notes
Donald Trump	Former President of the US
Hillary Clinton	Former US Secretary of State
Barack Obama	Former president of the US
Theresa May	Former prime minister of the UK
Emanuel Macron	President of France
Angela Merkel	Former Chancellor of Germany
Kim Jong Un	Supreme Leader of North Korea
Vladimir Putin	President of Russia

Table 2: Politicians selected for study 2

Figure 8: Participants' responses to questions "How would you describe your political outlook with regard to economic/social issues?" for Study 2.

To curate the stimuli for the present study, the tweets dataset from the previous study was filtered to tweets that included mentions of 8 different prominent politicians (see Table 2). Users were instructed that each set of tweets mentioning each politician were from a single source. To limit the impact of emotional text content on users' judgments, we downselected tweets with the following steps: we first conducted sentiment analysis on the tweet texts using Vader Sentiment (Hutto and Gilbert 2014). Next, for each set of tweets, we selected tweets with the highest neutral sentiment scores. Finally, we qualitatively evaluated and removed tweets with inaccurate scores from the sentiment analysis library. This resulted in 30 to 40 tweets mentioning each politician from mainstream news sources that were mostly of neutral tone. The images for this study were manipulated in a between-subjects manner such that users saw either happy, angry, or no images of each politician. For example, a user in the Happy condition, viewed tweets mentioning Kim Jung Un accompanied by happy images of him, while a user in the Angry condition evaluated the same tweets but with angry images of him, and the control (no-image) condition viewed no images (see Figure 7 for one of the example tweets).

The procedures of the study were equivalent to Study 1, with one exception: To measure the interaction between users' prior attitude towards each politician and the experimental conditions, for each set of tweets users first answered two questions in a pop-up form about their familiarity and favorability towards that politician on 5-point Likert scales.

Hypotheses

Since study 2 is a continuation of study 1, we expected to see a similar effect of angry emotions on content bias. We also hypothesized that users in the angry condition would rate sources as less credible. Guided by research on motivated reasoning (Kahan 2012), we expected to observe an interaction between users' prior favorability of different politicians and sources' negative and positive visual bias towards those politicians. More specifically, we hypothesized that favorability negatively interacts with angry emotion and positively interacts with happy emotion to predict users perceived bias and credibility. Furthermore, we also investigate the impact of users' familiarity with each politician on their judgments.

Dependent & Independent Variables

The dependent variables in Study 2 are identical to Study 1. We considered four total dependent variables(DV): (1)
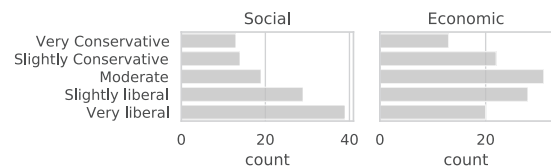
the tweet bias choice (bounded value between [0,1]), (2) uncertainty range around tweet bias (bounded value between [0,1]), (3) source credibility choice (bounded value between [0,1]), (4) uncertainty around source credibility choice (bounded value between [0,1]). For our independent variables (IV), we included the image emotion condition (angry, happy, or no image), as well as users' prior favorability and familiarity towards each politician.

Model Specification

For each model, we included users' unique IDs and the politician's name as random effects. After comparing multiple model specifications using AIC, we also included interaction terms between users' favorability and familiarity of each politician with the image emotion. The omitted reference conditions are image emotion = no image.

Participants

In study 2, we recruited a total of 126 participants. The average age of participants was 35 years old (std of 11.3). 81 participants were recruited from Amazon Mechanical Turk and received a 2-dollar incentive. To get closer to our pre-registration target, the rest of the participants were university students who received either research or course credit for their participation. Per our pre-registration, we excluded responses from 12 participants with missing responses (due to unexpected technical difficulties) resulting in 114 accepted responses (54 women, 59 men, and 1 prefer not to say; 87 white, 10 African American, 7 East Asian, 5 Hispanic, 3 Other Asian and 2 middle eastern). Participants were mostly liberal in regards to social, and political issues and more balanced in regards to economic issues (See Fig 8). 34 participants were randomly assigned to the angry condition, 42 participants were assigned to the happy condition, and the remaining 38 were assigned to the no image condition. Participants took an average of 21 minutes to complete the study.

Results

Content Bias. We used two mixed-effects beta regressions to study the effects of experimental conditions on users' judgment and uncertainty of bias of individual tweets (see fig 9-left). We found that users viewing tweets with angry images rated tweets as more biased in comparison to when no images are shown to users ($\beta = 0.511$ [0.18, 0.84], $z = 3.031, p < .01$). This is in line with our hypothesis that when the content of messages is the same, angry imagery leads to an increase in perceived bias. Similar to study 1, we did not
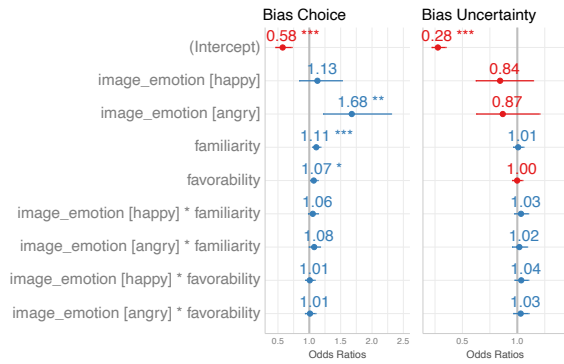
Figure 9: Study 2 fixed effects Odds Ratios for bias choice (left) and bias uncertainty (right). Error bars indicate 95% confidence intervals. Asterisks indicate statistical significance than zero using p-values: *** 99.9%, ** 99%, * 95%.
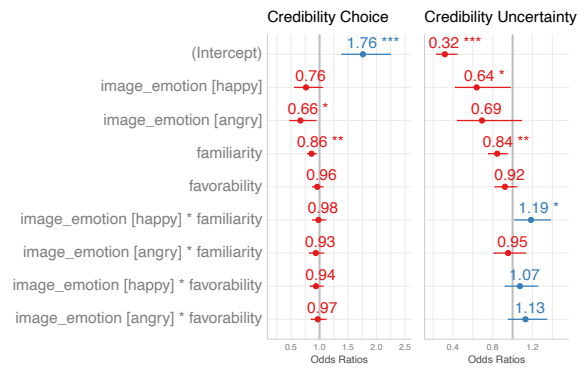

Figure 10: Study 2 fixed effects Odds Ratios for credibility choice (left) and credibility uncertainty (right). Error bars indicate 95% confidence intervals. Asterisks indicate statistical significance than zero using p-values: *** 99.9%, ** 99%, * 95%.

observe any effect in the happy condition. Moreover, we did not find any evidence of an interaction between favorability/familiarity and image emotion. We did, however, observe an overall positive effect of familiarity on users' perceived bias ($\beta = 0.115$ $[0.05, 0.17], z = 3.82, p < .001$). We also observed a small positive effect of favorability on perceived bias ($\beta = 0.074$ $[0.01, 0.13], z = 2.471, p < .05$). We did not observe any significant effects of any of the independent variables on users' uncertainty (See Fig 9).

Source Credibility: We used mixed-effects beta regression to investigate the effects of study conditions on users' judgments about source credibility (see Fig 10-left). We found that users rated sources as less credible when tweets were accompanied by angry facial expressions ($\beta = -0.397$ $[-0.73, -0.05], z = -2.308, p < .05$). This finding aligns with our hypothesis that angry facial expressions might lead to a decrease in the perceived source credibility. We also find that familiarity also decreases users' perceived credibility of sources ($\beta = -0.153$ $[-0.245, -0.06], z = -3.256, p < .01$). We did not observe an interaction effect between favorability and the image emotion.

For the mixed effect model of users' uncertainty, we observed a significant effect of happy emotion on users' judgment uncertainty ($\beta = -0.44$ $[-0.8 - 0.02], z = -2.09, p < .01$). We also observed that higher familiarity with subjects leads to more certain judgments ($\beta = -0.168$ $[-0.28 - 0.05], z = -2.9, p < .01$). We also observed a positive interaction between happy emotion and users' familiarity ($\beta = 0.16$ $[0.0240.31], z = 2.224, p < .05$), such that there was greater uncertainty in the happy condition for more familiar subjects.

Qualitative Analysis of Users' Comments

In study 2, we collected a total of 881 comments from 116 users. Similar to study 1, we analyzed users' descriptions of the rationale behind their judgments using NMF topic modeling (Cichocki and Phan 2009) and thematic analysis of the documents most representative of each topic. This

helped us categorize the 20 extracted topics into 5 higher-order themes. Four of the themes were similar to the ones we extracted from study 1 and discussed cues found in the text of the tweets. Here again, we will offer a qualitative overview of users' comments about images.

We identified three topics that contain descriptions from users related to visual information. Topic 19 with 71 comments from 34 users and Topic 14 with 47 comments from 24 users include mentions of facial expressions, angry emotions, and unflattering portrayals. For example, a user found the text of tweets mostly unbiased and explained how portrayed facial expressions of Hillary Clinton was the basis for their judgment: *"While the text didn't involve much biased words, the usage of certain images of Hillary Clinton depicting her facial expressions in an array of negative emotions showed a biased view."*

Some users also found a combination of text with "weirdly close up" and "unflattering" images leading them to believe a source is less credible: *"The texts were mostly bland, except for "...what do you think", which is a tabloid-like phrase for me. Photos were weirdly close-up facial views that were generally unflattering, which makes me wonder a bit about credibility..."*. Another interesting set of comments was about users perceiving the images as not correlating with the tweets: *"Most of the tweets were very normal, but there were a couple that had angry Macron pictures that did not correlate with the headlines presented."* And *"The majority of tweets were unbiased with their headlines. Some of the pictures might have been a bit questionable."*

A group of comments described how images did not influence users' judgments. Topic 16 with 66 comments from 32 users, includes examples of such commentary. For example, for tweets about Kim Jung Un, a user mentioned: *"The images were not the best images nor were they the worst images of him, but there was still a negative bias."* Another user mentioned how images and tweets related to Donald Trump were both neutral and honest looking: *"The tweets seemed to be honest and state the honest news about what is*
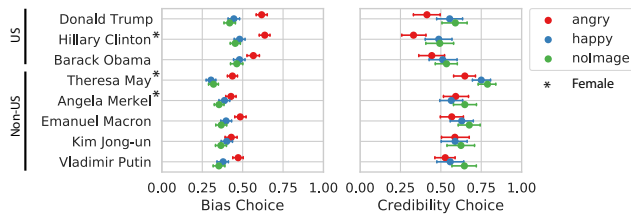
477

Figure 11: Mean and 95% bootstrapped confidence interval for bias and credibility choice for each set of tweets about each politician.

Figure 12: Favorability and familiarity counts (Study 2).

happening while the images associated with them."

Study 2 Discussion

We asked participants to judge sources that each focused on a specific politician. The text was constant for all users, while we manipulated the tweets to include either negative (angry) or positive (happy) facial expressions of those politicians or to include no images as a control condition. We found that in comparison to users in the no image condition, users in the angry condition found the content to be more biased and the sources to be less credible. We did not find a reverse effect of happy images on users' judgments. The difference between the happy and angry condition in our study could be better explained by a study on perceptions of negative or positive portrayals of politicians, in which Lubinger and Brantner found that participants mostly agreed on what constituted as negative, but perceptions of positive portrayals varied widely (Lobinger and Brantner 2015). This suggests that negative portrayals might be more salient and more likely to be commonly agreed upon and thus might have a stronger effect on users' judgments. Additionally, users' comments also included many mentions of angry, negative, or unflattering portrayals of these politicians, while we did not observe mentions of positive bias in users' comments.

Although the effect is rather small, we observed that users rate tweets about politicians they are more familiar with as more biased and to come from less credible sources (See Fig 12 for counts of favorability and familiarity responses for each politician). Recent work on judgments on misinformation suggests that prior exposure and familiarity with misinformation increases the perceived accuracy of content in which users would have to rely on their memory to assess the accuracy of news headlines or articles (Pennycook, Cannon, and Rand 2018). Since to assess the credibility and bias of anonymous content, users would not rely on their memories, we suspect that the impact of familiarity on these judgments might be different. Our comment analysis highlighted that users often rely on more analytical approaches and different heuristics such as negativity, word usage, or emotions in facial expressions to judge bias and credibility of sources. An explanation for the effect of familiarity on users' judgment about bias and credibility could be that when users are more familiar with politicians, they might be more sensitive to the details of texts and images of content they view and therefore possibly more likely to identify cues impacting their credibility and bias judgments.
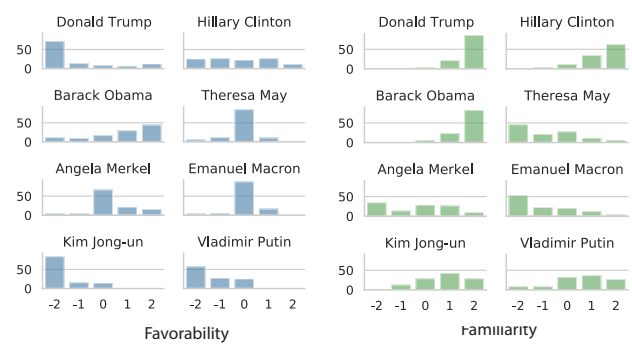
Finally, we observed a small overall effect of favorability on users' perception of tweet bias, but we did not observe an interaction between favorability and emotions in images. Although we elicited users' self-described political orientation, we elected to instead use favorability and familiarity of specific politicians to more directly study the impact of users' existing attitudes on their credibility and bias judgments. We expected that if a user is engaged in motivated reasoning, they would find attitude-consistent portrayals (e.g., an angry portrayal of a politician they dislike) as a cue for credibility, while portrayals that conflict with preexisting attitudes would be associated with lower credibility judgments. However, our results do not provide evidence that, in such experimental settings, editemotions in facial expressions impacts users who might be engaged in motivated reasoning based on their favorability towards politicians. This result is in line with work showing that motivated reasoning is not a primary factor in users' judgments when they are asked to evaluate news content (Pennycook and Rand 2019).

General Discussion and Conclusion

Study 1 showed that angry facial emotions lead to higher ratings of content bias, but not lower credibility ratings for sources. However, these judgments were possibly impacted by differences in the textual content and topics across the sources we used. In study 2 we controlled for the textual content while investigating the systematic negative or positive visual portrayals of politicians on users' perceptions of credibility and bias. Our results provide partial evidence for our hypothesis: systematic negative visual treatments of politicians led to both higher perceived content bias and lower perceived source credibility. However, we did not find strong evidence for our hypotheses around the interactions between users' prior favorability of politicians and sources' systematic negative or positive visual portrayal of those politicians.

We also explored the impact of different study conditions on users' certainty around their judgments. In study 1, misinformation sources were associated with more certain credibility judgments. One possible interpretation for this effect is that many misinformation sources are more likely to use content with more suspicious language and images (Volkova et al. 2019, 2017). In study 2, we observed that users were more certain in their credibility judgments when they were

more familiar with a politician. Users with greater direct knowledge of a politician may have been more sensitive to the accuracy of the content and were therefore able to make more confident credibility judgments.

Although these results provide evidence of the impact of emotional facial expressions in images on users' judgments, we found that users also rely heavily on other cues in the tweet content such as political orientation and opinionated language. It is also important to note that there are many more aspects related to emotions in images that remain to be studied. A comparison between angry, happy, and neutral emotions is an appropriate comparison which we omitted due to the difficulty of curating a balanced dataset with neutral images of all eight politicians. Moreover, within each of the angry or happy emotional categories, there are finer distinctions between different emotional facial expressions, ranging from extremely angry/happy to subtle frown/grin that might impact users' judgments. Comments about satirical images point to the importance of how subtle changes in facial expressions could impact users' perceptions of emotional facial expressions. There are also other emotional dimensions such as sadness, surprise, fear, or disgust that might potentially impact users' judgments. Many of these subtle differences might be present in our stimuli and may have impacted participants' judgments. Finally, facial expressions rarely contain one unique emotion, and subtle changes might communicate different meanings to individuals. In addition to facial expressions, prior research has established the influence of other visual features (including body expression and scene information) (Kret et al. 2013; Kret and Gelder 2012, 2013) on the perception of emotions. Although these factors were outside of the scope of the two studies reported in this paper, they are promising next steps to investigate. In future studies, Generative Adversarial Neural (GAN) networks (Goodfellow et al. 2020) can be used to produce image datasets with finer control over the facial expressions or other emotional content (Todorov et al. 2021).

We acknowledge some general limitations in the design and execution of our studies. First, participants in our study 1 were leaning liberal in regards to social issues and were of younger age, it is important to replicate these findings for a more balanced age and political orientation sample as well. Second, to remove potential confounds introduced by body posture and image backgrounds, we cropped all images to include only faces. However, this decision might have a generally negative effect on users' bias judgments as seen in some of the participants' comments and prior research identifying face size as a specific type of visual bias (Peng 2018). Additionally, for users' judgments, we did not measure differences between positive and negative content bias. Eliciting valenced judgments about content bias may provide more insight into how these judgments depend on users' pre-existing attitudes or familiarity with a subject. In the future, it is important to control for and study the impact of factors such as face size and positive/negative bias on users' judgments. Moreover, To limit the impact of users' preconceived notions of sources, we masked all account names from our stimuli. Even though encountering new and unknown sources is a realistic scenario and espe-

cially important in the context of misinformation, in many cases, users might have a self-selected set of sources they trust and refer to. It is possible that users' prior knowledge of sources significantly impacts how they evaluate content. For example, a user who trusts Fox News or Washington Post might be less sensitive to any visual bias portrayed by these sources toward a politician. An important future step for our research is to investigate how users update their trust in familiar sources when they observe a systematic usage of biased visual content. Other factors such as gender and political orientation of politicians are also likely to impact users' judgments of bias and credibility and require closer attention in future studies. Although we did not explicitly preregister and control for such factors in our experiments, we did not observe noticeable differences between politicians in our stimuli (See Fig 11).

Broader Perspective and Ethics

This study was conducted in experimental settings. The procedures of the study were approved by our institution's IRB, every participant signed an informed consent, and throughout the study, the participant's identities remained anonymous. We believe our results could potentially be used by malicious actors to further influence consumers' judgments. However, as most social media interactions are driven by polarized outrage (Rathje, Van Bavel, and van der Linden 2021), these results can engage the community in an ethical discourse on the curation of sensationalized visual media by mainstream and misinformation organizations. Moreover, to learn about the true impacts of content on users' beliefs and attitudes, we believe that results of experimental settings should be repeated "in the wild" (Mosleh, Pennycook, and Rand 2022). Such in-the-wild experiments, however, require attention to both platform rules (in our case, Twitter), and ethical guidelines for interacting with users on social media. These ethical questions for social media experiments are critical, as it is difficult to maintain anonymity on these platforms. In the future, as we move towards combating harmful information on social media, such ethical considerations become truly essential.

Conclusion

Across two consecutive preregistered studies with a total of 207 participants, we find evidence that negative (angry) facial expressions in social media news images lead to an increase in users' perception of bias in content. We also found that users judge sources to be less credible when they observe a potentially systematic negative portrayal of different politicians. While this paper was motivated by the prevalence and impact of misinformation on our democracies and societies, our current globally politicized and polarized political ecosystem calls for a more critical view of the whole media landscape. This paper highlights the importance of moving beyond the accuracy of textual content in the constant struggle to build trust in credible news sources. Sensationalized news content, even in subtle ways such as emotional facial expressions, might impair users' trust in otherwise trustworthy news. These effects can lead to harm-

ful outcomes evident in many current issues such as vaccine hesitancy and polarization.

References

Andalibi, N.; and Buss, J. 2020. The human in emotion recognition on social media: Attitudes, outcomes, risks. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16.

Arsenault, A.; and Castells, M. 2006. Conquering the minds, conquering Iraq: The social production of misinformation in the United States–a case study. *Information, Communication & Society*, 9(3): 284–307.

Berger, J.; and Milkman, K. L. 2012. What makes online content viral? *Journal of marketing research*, 49(2): 192–205.

Bowman, N. D.; and Cohen, E. 2020. 17 Mental Shortcuts, Emotion, and Social Rewards: The Challenges of Detecting and Resisting Fake News. *Fake News: Understanding Media and Misinformation in the Digital Age*, 223.

Campellone, T. R.; and Kring, A. M. 2013. Who do you trust? The impact of facial emotion and behaviour on decision making. *Cognition & emotion*, 27(4): 603–620.

Cichocki, A.; and Phan, A.-H. 2009. Fast local algorithms for large scale nonnegative matrix and tensor factorizations. *IEICE transactions on fundamentals of electronics, communications and computer sciences*, 92(3): 708–721.

de Melo, C.; Gratch, J.; Carnevale, P.; and Read, S. 2012. Reverse appraisal: The importance of appraisals for the effect of emotion displays on people's decision making in a social dilemma. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 34.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2020. Generative adversarial networks. *Communications of the ACM*, 63(11): 139–144.

Grabe, M. E.; and Bucy, E. P. 2009. *Image bite politics: News and the visual framing of elections*. Oxford University Press.

Hovland, C. I.; Janis, I. L.; and Kelley, H. H. 1953. *Communication and persuasion*. Yale University Press.

Hutto, C.; and Gilbert, E. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8.

Kahan, D. M. 2012. Ideology, motivated reasoning, and cognitive reflection: An experimental study. *Judgment and Decision making*, 8: 407–24.

Karduni, A.; Cho, I.; Wesslen, R.; Santhanam, S.; Volkova, S.; Arendt, D. L.; Shaikh, S.; and Dou, W. 2019. Vulnerable to misinformation? Verifi! In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 312–323.

Karduni, A.; Markant, D.; Wesslen, R.; and Dou, W. 2020. A Bayesian cognition approach for belief updating of correlation judgement through uncertainty visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 1–1.

Karduni, A.; Wesslen, R.; Santhanam, S.; Cho, I.; Volkova, S.; Arendt, D.; Shaikh, S.; and Dou, W. 2018. Can You Verifi This? Studying Uncertainty and Decision-Making About Misinformation using Visual Analytics. In *International Conference on Web and Social Media (ICWSM)*.

Knobloch, S.; Hastall, M.; Zillmann, D.; and Callison, C. 2003. Imagery effects on the selective reading of Internet newsmagazines. *Communication Research*, 30(1): 3–29.

Kret, M.; and Gelder, B. 2012. Islamic Headdress Influences How Emotion is Recognized from the Eyes. *Frontiers in psychology*, 3: 110.

Kret, M.; and Gelder, B. 2013. When a smile becomes a fist: The perception of facial and bodily expressions of emotion in violent offenders. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, 228.

Kret, M.; Roelofs, K.; Stekelenburg, J.; and de Gelder, B. 2013. Emotional signals from faces, bodies and scenes influence observers' face expressions, fixations and pupil-size. *Frontiers in Human Neuroscience*, 7.

Lazer, D. M.; Baum, M. A.; Benkler, Y.; Berinsky, A. J.; Greenhill, K. M.; Menczer, F.; Metzger, M. J.; Nyhan, B.; Pennycook, G.; Rothschild, D.; et al. 2018. The science of fake news. *Science*, 359(6380): 1094–1096.

Lo, W. H.; and Cheng, B. K. L. 2017. The use of melodramatic animation in news, presence and news credibility: a path model. *Journalism Studies*, 18(6): 787–805.

Lobinger, K.; and Brantner, C. 2015. Likable, funny or ridiculous? A Q-sort study on audience perceptions of visual portrayals of politicians. *Visual Communication*, 14(1): 15–40.

Martel, C.; Pennycook, G.; and Rand, D. G. 2020. Reliance on emotion promotes belief in fake news. *Cognitive research: principles and implications*, 5(1): 1–20.

Masch, L.; and Gabriel, O. W. 2020. How Emotional Displays of Political Leaders Shape Citizen Attitudes: The Case of German Chancellor Angela Merkel. *German Politics*, 29(2): 158–179.

McCabe, D. P.; and Castel, A. D. 2008. Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition*, 107(1): 343–352.

McInnes, L.; Healy, J.; and Astels, S. 2017. hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 2(11): 205.

Messaris, P.; and Abraham, L. 2001. The role of images in framing news stories. *Framing public life: Perspectives on media and our understanding of the social world*, 215–226.

Morris, M. R.; Counts, S.; Roseway, A.; Hoff, A.; and Schwarz, J. 2012. Tweeting is believing? Understanding microblog credibility perceptions. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*, 441–450.

Mosleh, M.; Pennycook, G.; and Rand, D. G. 2022. Field experiments on social media. *Current Directions in Psychological Science*, 31(1): 69–75.

Mullen, B.; Futrell, D.; Stairs, D.; Tice, D. M.; Baumeister, R. F.; Dawson, K. E.; Riordan, C. A.; Radloff, C. E.;

Goethals, G. R.; Kennedy, J. G.; et al. 1986. Newcasters' facial expressions and voting behavior of viewers: Can a smile elect a president? *Journal of Personality and Social Psychology*, 51(2): 291.

Pehlivanoglu, D.; Lin, T.; Deceus, F.; Heemskerk, A.; Ebner, N. C.; and Cahill, B. S. 2021. The role of analytical reasoning and source credibility on the evaluation of real and fake full-length news articles. *Cognitive research: principles and implications*, 6(1): 1–12.

Peng, Y. 2018. Same candidates, different faces: Uncovering media bias in visual portrayals of presidential candidates with computer vision. *Journal of Communication*, 68(5): 920–941.

Pennycook, G.; Cannon, T. D.; and Rand, D. G. 2018. Prior exposure increases perceived accuracy of fake news. *Journal of experimental psychology: general*, 147(12): 1865.

Pennycook, G.; and Rand, D. G. 2019. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188: 39–50.

Pennycook, G.; and Rand, D. G. 2021. The psychology of fake news. *Trends in cognitive sciences*.

Petty, R. E.; Wegener, D. T.; and Fabrigar, L. R. 1997. Attitudes and attitude change. *Annual review of psychology*, 48(1): 609–647.

Pornpitakpan, C. 2004. The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of applied social psychology*, 34(2): 243–281.

Porter, S.; Bellhouse, S.; McDougall, A.; Ten Brinke, L.; and Wilson, K. 2010. A prospective investigation of the vulnerability of memory for positive and negative emotional scenes to the misinformation effect. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 42(1): 55.

Rathje, S.; Van Bavel, J. J.; and van der Linden, S. 2021. Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, 118(26).

Reis, J.; Benevenuto, F.; de Melo, P. O.; Prates, R.; Kwak, H.; and An, J. 2015. Breaking the news: First impressions matter on online news. *arXiv preprint arXiv:1503.07921*.

Richards, B. 2007. *Emotional governance: Politics, media and terror*. Springer.

Schroff, F.; Kalenichenko, D.; and Philbin, J. 2015. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 815–823.

Schweitzer, N.; Baker, D. A.; and Risko, E. F. 2013. Fooled by the brain: Re-examining the influence of neuroimages. *Cognition*, 129(3): 501–511.

Shi, Z.; Wang, A.-L.; Emery, L. F.; Sheerin, K. M.; and Romer, D. 2017. The importance of relevant emotional arousal in the efficacy of pictorial health warnings for cigarettes. *Nicotine & Tobacco Research*, 19(6): 750–755.

Small, D. A.; and Verrochi, N. M. 2009. The face of need: Facial emotion expression on charity advertisements. *Journal of marketing research*, 46(6): 777–787.

Todorov, A. T.; Uddenberg, S. D.; Peterson, J. C.; Griffiths, T. L.; and Suchow, J. W. 2021. Data-driven, photorealistic social face-trait encoding, prediction, and manipulation using deep neural networks. US Patent App. 17/026,797.

Tormala, Z. L.; and Petty, R. E. 2004. Source credibility and attitude certainty: A metacognitive analysis of resistance to persuasion. *Journal of Consumer Psychology*, 14(4): 427–442.

Tormala, Z. L.; and Rucker, D. D. 2007. Attitude certainty: A review of past findings and emerging perspectives. *Social and Personality Psychology Compass*, 1(1): 469–492.

Van Kleef, G. A.; De Dreu, C. K.; and Manstead, A. S. 2010. An interpersonal approach to emotion in social decision making: The emotions as social information model. In *Advances in experimental social psychology*, volume 42, 45–96. Elsevier.

Vlasceanu, M.; Goebel, J.; and Coman, A. 2020. The Emotion-Induced Belief Amplification Effect. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Volkova, S.; Ayton, E.; Arendt, D. L.; Huang, Z.; and Hutchinson, B. 2019. Explaining multimodal deceptive news prediction models. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, 659–662.

Volkova, S.; Shaffer, K.; Jang, J. Y.; and Hodas, N. 2017. Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, 647–653.

Wallace, L. E.; Wegener, D. T.; and Petty, R. E. 2020a. Influences of source bias that differ from source untrustworthiness: When flip-flopping is more and less surprising. *Journal of personality and social psychology*, 118(4): 603.

Wallace, L. E.; Wegener, D. T.; and Petty, R. E. 2020b. When sources honestly provide their biased opinion: Bias as a distinct source perception with independent effects on credibility and persuasion. *Personality and Social Psychology Bulletin*, 46(3): 439–453.

Wardle, C.; and Derakhshan, H. 2017. Information Disorder: Toward an interdisciplinary framework for research and policymaking. *Council of Europe report, DGI (2017)*, 9.

Wobbrock, J. O.; Hattatoglu, L.; Hsu, A. K.; Burger, M. A.; and Magee, M. J. 2021. The Goldilocks zone: young adults' credibility perceptions of online news articles based on visual appearance. *New Review of Hypermedia and Multimedia*, 1–46.

Zillmann, D.; Knobloch, S.; and Yu, H.-s. 2001. Effects of photographs on the selective reading of news reports. *Media Psychology*, 3(4): 301–324.