Article

new media & society
2019, Vol. 21(2) 438–463

DOI: 10.1177/1461444818799526
journals.sagepub.com/home/nms
SAGE

Fake images: The effects of source, intermediary, and digital media literacy on contextual assessment of image credibility online

Cuihua Shen
University of California, Davis, USA

Mona Kasra
University of Virginia, USA

Wenjing Pan, Grace A Bassett and Yining Malloch
University of California, Davis, USA

James F O'Brien
University of California, Berkeley, USA

Abstract

Fake or manipulated images propagated through the Web and social media have the capacity to deceive, emotionally distress, and influence public opinions and actions. Yet few studies have examined how individuals evaluate the authenticity of images that accompany online stories. This article details a 6-batch large-scale online experiment using Amazon Mechanical Turk that probes how people evaluate image credibility across online platforms. In each batch, participants were randomly assigned to 1 of 28 news-source mockups featuring a forged image, and they evaluated the credibility of the images based on several features. We found that participants' Internet skills, photo-editing experience, and social media use were significant predictors of image

Corresponding author:
Cuihua Shen, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA.
Email: cuishen@ucdavis.edu

credibility evaluation, while most social and heuristic cues of online credibility (e.g. source trustworthiness, bandwagon, intermediary trustworthiness) had no significant impact. Viewers' attitude toward a depicted issue also positively influenced their credibility evaluation.

Keywords
Digital media literacy, image credibility, image manipulation, intermediary, online images, source credibility

The ubiquitous availability of easy-to-use software for editing digital images brought about by rapid technological advances of the 21st century has dramatically decreased the time, cost, effort, and skill required to fabricate convincing visual forgeries. Often distributed through trusted sources such as mass media outlets, perhaps unknowingly, these manipulated images propagate across social media with growing frequency and sophistication. Moreover, the technology that allows for manipulating or generating realistic appearing images has far outpaced the technological development of methods for detecting fake imagery and even experts often cannot rely on visual inspection to distinguish authentic digital images from forgeries. Bad actors can thus easily publish manipulated visual content to deceive their viewers, inflicting cognitive stress, exploiting prior beliefs, or influencing individuals' decisions and actions.

Although it is difficult to say how prevalent undetected occurrences of fake imagery are, numerous examples have been exposed in which manipulated images have caused substantial harms at individual, organizational, and societal levels. For instance, an image of Senator John Kerry and Jane Fonda sharing the stage at a Vietnam era antiwar rally emerged during the 2004 presidential primaries as Senator Kerry was campaigning for the Democratic nomination. The accompanying caption stated, "Actress and Anti-War Activist Jane Fonda Speaks to a crowd of Vietnam Veterans as Activist and Former Vietnam Vet John Kerry (LEFT) listens and prepares to speak next concerning the war in Vietnam (AP Photo)." The forged photograph, however, was created by compositing together two separate photos that separately depicted Kerry and Fonda. The edited image showing them together gave the false impression that Kerry shared the controversial antiwar views of activist Jane Fonda (Light, 2004). In a more recent example, in January 2014, the Associated Press news agency fired its Pulitzer prize-winning photographer Narciso Contreras for digitally removing an object from one of his widely distributed photographs of the Syrian civil war (*The Guardian*, 2014). This case has stirred an ongoing and contested discussion about the authenticity of digital photographs, the potential repercussions of image manipulation, and the ethics code in photojournalism. Numerous other examples exist where fake imagery has been used to distort the truth and manipulate viewers (For more examples, see http://pth.izitru.com/). It is unclear how prevalent are instances of undetected photo manipulation.

The damage done by manipulated imagery is real, substantial, and persistent. Studies suggest that manipulated images can distort viewer's memory (Wade et al., 2002)—therefore further enhancing the credibility of these images—and even influence decision-making behaviors such as voting (Bailenson et al., 2008; Nash et al., 2009).

Moreover, even when individuals do become aware of the true nature of a forgery, the harmful impact of misinformation on their perception, memory, emotions, viewpoints, and attitude toward a topic can linger (Sacchi et al., 2007). Quite often the distribution of fake images will far surpass the distribution of any correction or attempt to expose the forgery (Friggeri et al., 2014). The factors combine to make image manipulation an extremely effective and difficult to combat manipulation method.

While there is a growing awareness that images should no longer be automatically assumed to be credible, authentic, or reliable sources of information, the general public remains vulnerable to visual deception. Due to the scope and speed of information dissemination across social media websites, the potential for ill-intentioned players to inflict emotional distress or to purposefully influence opinions, attitudes, and actions through visual misinformation poses a severe and growing societal risk. Yet we know distressingly little about how online viewers assess or make credibility judgments of online images. This article details a large-scale online experiment of image credibility that seeks to understand how individuals evaluate manipulated images that accompany online stories, and what features (image-related and non-image-related) impact their credibility judgment. The images tested in this study were altered using common manipulation techniques: composition, elimination, and retouching (identified in Kasra et al., 2018).

The research design was informed by earlier research on social and heuristic approaches to credibility judgment as well as by our previous exploratory findings on online image credibility (Kasra et al., 2018). Previous research in this area has either predominantly focused on fake image detection using machine learning approaches (Gupta et al., 2013), or on the credibility of textual information, such as websites and blogs (Allcott and Gentzkow, 2017; Morris et al., 2012; Wineburg and McGrew, 2016). These studies tend to assume that individuals make credibility evaluations on their own without considering that decisions are heavily influenced by one's social networks. Our study is among the first to test the social and cognitive heuristics of information credibility and evaluation in the context of image authenticity.

Theory and hypotheses

Online information credibility

In their article examining the elements of computer credibility, Tseng and Fogg (1999) argue that credibility, or "believability," is not an objective quality of the information itself but ultimately a perception of the user. They identify the trustworthiness and perceived expertise of the source of information as the most important factors in forming credibility perceptions (Tseng and Fogg, 1999).

In recent years, instances of deliberate circulation of misinformation and fake news across online media have become commonplace. A study found that "the average American adult saw on the order of one or perhaps several fake news stories in the months around the [2016 US] election, with just over half of those who recalled seeing them believing them" (Allcott and Gentzkow, 2017: 211). Such a large volume of (mis) information in the contemporary media environment presents formidable cognitive challenges for credibility judgment. Numerous studies have consistently demonstrated that

users rarely perform any evaluation behaviors (such as seeking out other sources to validate information, or checking out the author) to verify the credibility of online information (Metzger, 2007). To save cognitive time and effort, they instead process and determine the credibility of the information by relying on social and heuristic cues (Fogg et al., 2003; Metzger et al., 2010). These strategies include social information pooling (especially from like-minded others), cognitive heuristics (e.g. website reputation, endorsements, consistency, and expectancy violation), and persuasive intent (Metzger et al., 2010).

Social network sites such as Facebook and Twitter introduce a host of system-generated metrics (e.g. number of followers) as cognitive heuristics for credibility judgment (Sundar, 2008; Westerman et al., 2012, 2014). Similarly, information sharing behaviors on social network sites and news media sites, facilitated by the prevalence of "social buttons" (Gerlitz and Helmond, 2013), present additional cognitive heuristics for credibility evaluation. For instance, encountering a "secondhand" Facebook post, originally shared by the New York Times but now reshared on an individual's or organization's page, could complicate users' credibility assessment. Unfortunately, there has been relatively little empirical evidence on how the credibility of intermediaries may affect people's credibility judgment, especially when the credibility of the intermediary is inconsistent with that of the original source. It is therefore crucial that, in addition to features related to the original source, research on credibility evaluation takes into account the intermediaries, or "second-order" sources that share or endorse information originally published by someone else.

Furthermore, most online credibility research to date focuses on textual information. Among the few empirical studies specifically on image credibility, for instance, Gupta et al. (2013) identified more than 10,000 tweets containing fake photos, and used machine learning models to compare fake images to real ones based on several features. These features ranged from characteristics of the source (e.g. number of followers and whether the user is verified) to the quality of the tweet (e.g. length, sentiment, or hashtags used). Their result showed that tweet-based features identified fake from real images very well, while source (user) based features performed poorly.

Prior research suggests that viewers tend to believe the content depicted in online images. A study measuring students' ability in evaluating online sources of information also found that most high school students accept photographs as facts without verifying them (Wineburg and McGrew, 2016). Similarly, an exploratory study based on groups of US college students found that, in general, users are overly trusting toward images on the web (Kasra et al., 2018). More importantly, this study also revealed that viewers made their credibility judgments based on non-image-related features such as source and media channel, instead of image-related features such as inconsistencies in lighting, shadow, or color (Kasra et al., 2018).

This article reports the results of a comprehensive experiment on contextual assessment of image credibility online, measuring several factors such as the effects of source, intermediary, and digital media literacy on viewers' contextual assessments. In the following sections, we present hypotheses on six major factors that may influence credibility evaluation: source trustworthiness, source and media type, intermediary, bandwagon cues, digital skills and experiences, and pro-issue attitude.

Source trustworthiness

Decades of credibility research concludes that the reputation of the source is an important credibility heuristic (Metzger et al., 2010), and that credibility lies foremost in the trustworthiness and expertise of the source itself (Tseng and Fogg, 1999). Users tend to transfer the reputation of the source (companies as well as news organizations) to the content itself (Metzger et al., 2010). In a previous focus group study on fake images, most participants relied heavily on the source of online information, such as nationally recognized news organizations, to determine the credibility of an image (Kasra et al., 2018). So, source trustworthiness appears to play a critical role in evaluating online images as well. Therefore,

H1: Images from more credible sources will be perceived as more credible than those from less credible sources.

Source and media type

With an abundance of technical affordances, the social web provides many opportunities for individuals to consume and disseminate information, turning individuals into important, independent sources of information. But the stature and influence of online sources vary greatly as each source may convey a different level of expertise and authority. For example, news organizations are likely perceived as having more expertise than an individual in producing online news stories and images (Flanagin and Metzger, 2007). In addition, where the information is published can also influence credibility. Metzger et al. (2010) found that some media platforms are viewed with more skepticism because of their open editing structure (e.g. Wikipedia). By allowing users to self-publish and share without a central gatekeeper, social media sites such as Facebook and Twitter evoke similar level of skepticism among Internet users (Kasra et al., 2018).

In this article, we purposefully combine media platform with the type of source into one factor, as the two are often conflated in the contemporary media environment (Metzger et al., 2003). For example, most news media publish their stories on multiple platforms to reach a larger audience and because they have the capacity to do so. In contrast, most information generated by individuals is circulated on social media sites only. Therefore, we propose the following hypotheses:

H2(a): Images from news organizations will be perceived as more credible than those from individuals.

H2(b): Images from a news organization's official website will be perceived as more credible than those from their social media accounts.

Intermediaries

Information sharing and curating behaviors online, fueled by the prevalence of "social buttons" on social network sites and news media websites (Gerlitz and Helmond, 2013),

foregrounds the importance of second-order sources, or *intermediaries*, that share or endorse information published originally by someone else. However, few studies have examined how people's credibility evaluation varies based on the perceived trustworthiness of intermediaries. Metzger et al. (2010) use the term "endorsement heuristic," or conferred credibility, to describe people's tendency to perceive information as credible if others think so, especially when the endorser is considered credible itself (e.g. a site is more credible if it is endorsed by *New York Times*). However, it is unclear if resharing an image by an intermediary, or endorser, regardless of their reputation, would have any direct effect on the credibility of the image itself. The potential discrepancy between the trustworthiness of the original source and the trustworthiness of the intermediary is also equally unexplored. Therefore, we propose the following hypothesis and research questions:

H3: Images from more credible intermediaries will be perceived as more credible than those from less credible intermediaries.

RQ1: Are images shared by an intermediary perceived as more credible than images without an intermediary?

RQ2: Do source credibility and intermediary credibility interact with each other in affecting viewers' image credibility evaluation?

Bandwagon

The social buttons (like, share, favorite, etc.) available on websites and social media often aggregate opinions about specific stories. Several studies have shown consistent evidence for "bandwagon" effects where people are more likely to agree with a perceived consensus from aggregate metrics (Lee and Sundar, 2013; Sundar, 2008). In asking participants to select and read news articles with different recommendation ratings, Knobloch-Westerwick et al. (2005) found that higher rated articles were selected more often. We expect the same heuristic for online images, therefore:

H4: Images with higher levels of bandwagon cues will be perceived as more credible than those with lower levels of bandwagon cues.

Skills and experiences

We now turn to individual characteristics as predictors of credibility judgment. Familiarity and expertise with various online platforms, or "digital media literacy," are found to inform users' credibility assessment (Choi and Stvilia, 2015). Internet skills generally refer to users' ability to use various online media effectively (Hargittai and Hsieh, 2012). Such skills enable users to correctly identify cues such as source, intermediary, and bandwagon in order to make credibility assessments (Choi and Stvilia, 2015). In the context of online images, we conjecture that in addition to general familiarity with Internet and social media, experiences and proficiency related to photography or digital imaging would be especially useful. Therefore,

H5: People with greater levels of (a) photography and digital imaging experience and (b) general Internet skills will be more likely to perceive fake images as less credible than people with less experience/skills.

H6: People who use (a) Facebook and (b) Twitter more often will perceive fake images as less credible than people who use (a) Facebook and (b) Twitter less frequently.

Issue attitudes

Finally, confirmation bias is a well-established finding in the context of credibility judgment (e.g. Knobloch-Westerwick et al., 2015), in that people are more likely to perceive something as credible if it confirms their existing beliefs and opinions. This confirmation bias is probably even more pronounced for information related to politics or current events (Metzger et al., 2010). Therefore,

H7. People who have a high level of pro-attitude toward the issue depicted in the image will perceive it as more credible than people with a low level of pro-attitude.

Methods

Study design

Based on previous findings reviewed above, we identified 4 factors to be manipulated and tested in the experiment and statistically controlled 2 factors. We adopted a partial factorial design: Source credibility (High Credibility/Low Credibility) Source and Media Type (Website/Social Media Organization Account/Social Media Individual Account) × Intermediary (No/High Trust/Low Trust) × Bandwagon (High/Low), resulting in 28 unique conditions for each image tested (Table 1).

Note that a fully factorial design would have included 36 conditions ($2 \times 3 \times 3 \times 2 = 36$) but our partial factorial design had only 28 conditions instead. We omitted 8 conditions where an intermediary shared content from a news organization's website (e.g. a person sharing a story on Facebook from the New York Times website), because such conditions typically result in a thumbnail of the original image, which is significantly smaller than the image presented on the original website. In such cases, readers are required to click through the link in order to view the image in its original form. To ensure all images were presented and consumed in the same size and resolution without further clicking (see next section for mockup creation), we decided to exclude these 8 conditions, while keeping the conditions where an intermediary is sharing content from a news organization's social media account (e.g. a person sharing a story on Facebook from the New York Times Facebook account).

Source trustworthiness. In order to select the news organizations and individuals as sources of various levels of trustworthiness in our study manipulation, we conducted a pretest using a list of news organizations tested in a Pew report on media trustworthiness (Mitchell et al., 2014), including the *New York Times*, CNN, Fox News, and Buzzfeed. Individual sources consisted of celebrities and public figures such as Bill Gates, Diane Sawyer, and Anderson Cooper, who are rated in a Reader's Digest poll of trustworthiness

Table 1. Experimental conditions.

Condition	Source trustworthiness	Source used	Source and media type	Intermediary	Intermediary used	Bandwagon	Sample bandwagon cues
1	High	New York Times	Org. Website	NA		Low	50 likes, 11 shares, 1000 page views
2	High	New York Times	Org. Website	NA		High	22,575 likes, 1490 shares, 87,501 page views
3	High	New York Times	Org. Twitter/ Facebook	NA		Low	5 favorites, 3 retweets
4	High	New York Times	Org. Twitter/ Facebook	NA		High	64,540 favorites, 26361 retweets
5	High	New York Times	Org. Twitter/ Facebook	Low	Buzzfeed	Low	5 favorites, 3 retweets
6	High	New York Times	Org. Twitter/ Facebook	Low	Buzzfeed	High	64,540 favorites, 26361 retweets
7	High	New York Times	Org. Twitter/ Facebook	High	NPR/Bill Gates	Low	5 favorites, 3 retweets
8	High	New York Times	Org. Twitter/ Facebook	High	NPR/Bill Gates	High	64,540 favorites, 26361 retweets
9	High	Bill Gates	Personal Twitter/ Facebook	NA		Low	3 favorites, 2 retweets
10	High	Bill Gates	Personal Twitter/ Facebook	NA		High	30,635 favorites, 10,000 retweets
11	High	Bill Gates	Personal Twitter/ Facebook	Low	Rachael Hughes/ Mark Smith	Low	3 favorites, 2 retweets
12	High	Bill Gates	Personal Twitter/ Facebook	Low	Rachael Hughes/ Mark Smith	High	30,635 favorites, 10,000 retweets
13	High	Bill Gates	Personal Twitter/ Facebook	High	NPR	Low	3 favorites, 2 retweets
14	High	Bill Gates	Personal Twitter/ Facebook	High	NPR	High	30,635 favorites, 10,000 retweets

Table 1. (Continued)

Condition	Source trustworthiness	Source used	Source and media type	Intermediary	Intermediary used	Bandwagon	Sample bandwagon cues
15	Low	Buzzfeed	Org. Website	NA		Low	50 likes, 11 shares, 1000 page views
16	Low	Buzzfeed	Org. Website	NA		High	22,575 likes, 1490 shares, 87,501 page views
17	Low	Buzzfeed	Org. Twitter/ Facebook	NA		Low	5 favorites, 3 retweets
18	Low	Buzzfeed	Org. Twitter/ Facebook	NA		High	64,540 favorites, 26,361 retweets
19	Low	Buzzfeed	Org. Twitter/ Facebook	Low	Rachael Hughes/ Mark Smith	Low	5 favorites, 3 retweets
20	Low	Buzzfeed	Org. Twitter/ Facebook	Low	Rachael Hughes/ Mark Smith	High	64,540 favorites, 26,361 retweets
21	Low	Buzzfeed	Org. Twitter/ Facebook	High	NPR/Bill Gates	Low	5 favorites, 3 retweets
22	Low	Buzzfeed	Org. Twitter/ Facebook	High	NPR/Bill Gates	High	64,540 favorites, 26,361 retweets
23	Low	Rachael Hughes/ Mark Smith	Personal Twitter/ Facebook	NA		Low	1 favorites, 1 retweets
24	Low	Rachael Hughes/ Mark Smith	Personal Twitter/ Facebook	NA		High	261 favorites, 90 retweets
25	Low	Rachael Hughes/ Mark Smith	Personal Twitter/ Facebook	Low	Buzzfeed	Low	1 favorites, 1 retweets
26	Low	Rachael Hughes/ Mark Smith	Personal Twitter/ Facebook	Low	Buzzfeed	High	261 favorites, 90 retweets
27	Low	Rachael Hughes/ Mark Smith	Personal Twitter/ Facebook	High	Bill Gates	Low	1 favorites, 1 retweets
28	Low	Rachael Hughes/ Mark Smith	Personal Twitter/ Facebook	High	Bill Gates	High	261 favorites, 90 retweets

(Smith, n.d.). We also included two generic individuals, Rachael Hughes and Mark Smith to represent ordinary social media users. 53 college undergraduate students in a US University were recruited to pretest different manipulations, in which various sources (including both news organizations and individuals) were presented to participants to obtain their ratings of source trustworthiness.

The pretest results revealed the *New York Times* as the most trustworthy among news organizations, so it was included in our study as a highly credible source. National Public Radio (NPR) also scored consistently high and was included as a highly credible intermediary. Buzzfeed was regarded as a low credibility news organization (Table 1). As for individuals, Bill Gates was rated the most trustworthy individual. Generic individuals Rachael Hughes and Mark Smith received low credibility ratings and were both included in the manipulation as less credible individuals to counterbalance possible biases due to gender.

Source and media type. This factor included three levels: a news organization's website (e.g. New York Times' website), a news organization's social media account (e.g. New York Times' Twitter/Facebook account), and an individual's social media account (e.g. Bill Gates' Twitter/Facebook account). We did not consider an individual's website as a separate media platform, as based on prior research this combination is rarely a source of news information (Flanagin and Metzger, 2007). To assess the respective impact of Twitter and Facebook, we adopted Twitter as the social media account for half of the images tested, and Facebook for the other half (see Table 1).

Intermediary. This factor included three levels: no intermediary, intermediary with low trustworthiness, and intermediary with high trustworthiness. We used the same or similar news sources to manipulate the intermediary who helps diffuse images from the original source on social media sites. Highly credible intermediaries included National Public Radio (news organization) and Bill Gates (individual), and less credible intermediaries included Buzzfeed (news organization) and Rachael Hughes/Mark Smith (individual).

Bandwagon. We manipulated the numbers of likes, shares, favorites, retweets, and/or page views according to the specific source/intermediary and media channel in each condition. In the above-mentioned pretest, we asked participants to indicate the number of bandwagon cues considered "very many" or "very few" for different platforms (Facebook vs Twitter) and different sources (news organizations vs individuals). Participants were asked questions such as "If a known public figure (e.g. Bill Gates) tweets on Twitter, I would consider the tweet to have very many (or few) retweets if it had ___ retweets." The results from the pretest were used to establish what numbers would be used in the actual experiment for signaling a high or low level of bandwagon cues. For example, 5 favorites and 3 retweets were considered "low" for the New York Times on Twitter, while 64,540 favorites and 26,361 retweets were considered "high."

Images and mockups

A total of 6 fake images were chosen to be included in the study (Figures 1 to 6). Our criteria for selecting these images were to offer a wide range of topics and common

Figure 1. Mockup for Condition #2: A fake news article purportedly from the *New York Times* website. The image was modified to depict a bridge collapse.

Figure 2. Mockup for Condition #21: A fake Facebook post allegedly created by Buzzfeed and shared by Bill Gates. The image was composed by layering three separate images.

image manipulation techniques and methods (elimination, retouching, and composition identified in Kasra et al., 2018). The images were fake in that the content of each composition was purposefully cropped, changed, combined, or repaired. Misattribution (presenting an untouched image in an unrelated context or under false pretense) was not tested in the current study. The content of these images ranged from political portrayal,

Figure 3. Mockup for Condition #18: A fake Twitter post allegedly created by Buzzfeed. The image was composed by layering two separate images to create a cat-mouse chimera.

Figure 4. Mockup for Condition #19: A Facebook post allegedly created by Buzzfeed and shared by a generic person, Mark Smith. The image was composed by layering three separate images to depict a school in Africa.

political event, scientific discovery, natural disaster, and social issues. The images depicted (1) a bridge collapse in China, (2) a gay couple accompanied by their children, 3) genetically modified mouse with a cat's head, 4) a school in Africa, (5) a bombing in Syria, and (6) a Hispanic politician meeting with students. A short textual description

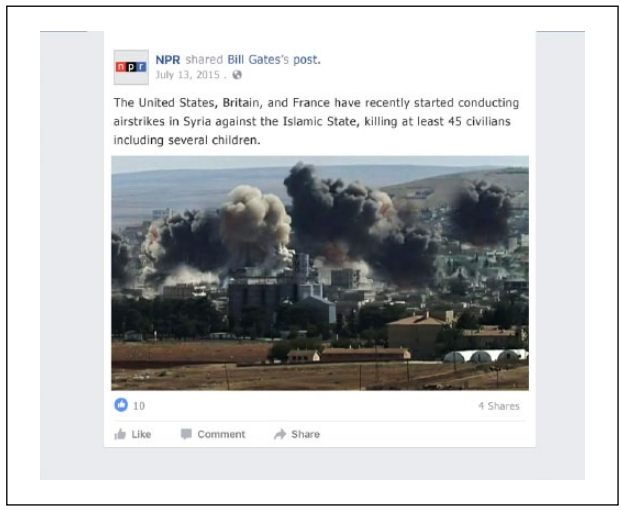

Figure 5. Mockup for Condition #13: A Facebook post allegedly created by Bill Gates and shared by NPR. The image was modified to display airstrikes in Syria.

Figure 6. Mockup for Condition #8: A fake Twitter post purportedly created by the *New York Times* and retweeted by Bill Gates. The image was composed by layering two separate images to show a politician meeting with students.

accompanied all conditions. The captions provided brief textual information about the content depicted, similar to how images are shared and viewed on the Web.

We created mockup compositions for each image based on the 28 experimental conditions, resulting in 28 mockups. The total number of mockups tested was 168 (28

conditions × 6 images). Every participant was exposed to only one mockup depicting one single image. For Figures 1, 3, and 6 (bridge collapse, genetically modified mouse, Hispanic politician), we used Twitter interface to create the mockups wherever appropriate. For Figures 2, 4, and 5 (gay couple, African school, and Syrian bombing), we used Facebook interface to create the mockups wherever appropriate. We used the actual social media handles and profile pictures for all news organizations and public figures. The two individuals, Rachael Hughes and Mark Smith, had generic Twitter and Facebook silhouettes instead of personalized profile pictures to keep them consistent and prevent partiality (see Figures 1 to 6). Note that silhouette profile images may imply anonymity which could adversely affect trustworthiness. However, Rachael Hughes and Mark Smith are intended to be low credibility individuals so reduced trustworthiness due to silhouette usage aligns with our study design.

Our goal was to understand (1) how viewers evaluate image credibility online and (2) what contextual cues and features (image-related and non-image-related) impact their credibility judgment. By including only fake images in the study instead of using unaltered, original images, we wanted to avoid the possibility of the participants having previously encountered one of the images. Such familiarity could have influenced their credibility evaluation without consideration of the features we attempted to test in this study.

Measures

Image credibility. Six items on a 7-point scale (1 = strongly disagree, 7 = strongly agree) were used to measure participants' perception of the credibility of the photo. The items were adapted from Flanagin and Metzger's (2007) scale of message credibility online and modified for use in the current study, and assessed the extent to which participants perceived the image to be believable, original, authentic, fake, manipulated, and retouched. Negatively worded items were reverse-coded. Then all items were averaged to create a composite credibility score (Cronbach's alpha = .95 for the whole sample).

Internet skills. Participants' Internet skills were measured by their self-reported familiarity with 10 Internet-related terms (e.g. phishing and tagging) on a 5-point Likert-type scale (Hargittai and Hsieh, 2012). A composite score of Internet skills was obtained by taking the mean of these 10 items (Cronbach's alpha = .92).

Digital imaging experiences and skills. Two items on a 5-point scale (1 = None, 5 = I'm an expert) were used to measure participants' digital imaging (e.g. photo editing) experiences and skills (Greer and Gosen, 2002; Cronbach's alpha = .74).

Facebook use/Twitter use. We first asked if participants had a Facebook account (Y/N). If they did, we proceeded with a six-item instrument on a 5-points Likert-type scale (1 = strongly disagree, 5 = strongly agree), adapted from (Ellison et al., 2007), to measure participants' intensity of Facebook use (e.g. "Facebook is part of my everyday activity"; "I would be sorry if Facebook shut down"). A composite intensity score was created by

taking the average of these 6 items (Cronbach's alpha = .90). People who did not have a Facebook account skipped this scale. Twitter use was measured in a similar manner (Cronbach's alpha = .91).

Pro-issue attitude. Two issue-relevant questions were used to measure participants' pre-existing attitudes toward the issue depicted in the image they were about to encounter in the mockup. The issue relevant questions were adapted from Treier and Hillygus (2009) and modified to fit each of the images tested. For example, the issue attitude question for Figure 3 (genetically modified mouse) asked the participant whether it is ethical or acceptable to genetically modify animals for research purposes. Negatively worded questions were reversed coded, and then the two items were averaged to create a composite score of pro-issue attitude.

Demographics. At the end of the survey, participants were asked to indicate their biological sex, age, race, annual household income, and education level. Participants' age and sex were included in our analysis as control variables.

Participants and procedure

Participants were recruited from Mechanical Turk (www.mturk.com, MTurk hereafter) and then participated in a survey hosted on Qualtrics. Participants who were younger than 18 years and/or not living in the United States at the time of the study were excluded, because some survey questions depended on basic knowledge of political issues relevant for the US audience. The experiment was conducted for 6 consecutive batches, each batch featuring one fake image in 28 conditions/mockups. Participants who had previously responded to one batch of data collection were excluded from the subsequent batches.

For batches #1 and #2 featuring a bridge collapse in China and a gay couple with their children, we recruited at least 30 participants per condition, resulting in around 840 participants per batch. Preliminary analysis yielded consistent results in both studies (more in the Results section). Since participants were only allowed to take part in one of our 6-batch series, the eligible MTurk pool became smaller as we proceeded with each subsequent study and recruitment process became longer (see Stewart et al., 2015: for an estimate of MTurk worker pool size). Therefore, for all following batches (Figures 3 to 6), we recruited at least 15 participants per condition, resulting in at least 420 participants per batch. There were 3476 total participants in the 6-batch experiment.

Participants responding to the request on MTurk first read the consent form and confirmed to proceed. They were then randomly assigned to 1 of the 28 image conditions/mockups. To ensure that the images were shown consistently with the same dimensions across the participant pool, we instructed participants who were detected to be on a mobile device to switch to a desktop or laptop computer; otherwise, they were not permitted to proceed with the survey. Participants first answered questions regarding Internet skills, digital imaging experience, and pro-issue attitude in random order. They were then presented with a mockup and instructed to look at the image carefully and rate its credibility (item order randomized). To make sure that participants spent enough time

examining the image before answering the credibility questions, the "Next" button was not displayed on the page until 30 seconds later. After rating image credibility, participants answered questions on their social media use, followed by demographic questions. The entire survey took approximately 5 minutes to complete. After completion, participants were paid $0.25 for the task.

In order to compare how the MTurk sample may deviate from typical undergraduate samples often employed in social science studies, including web credibility studies (Hargittai and Hsieh, 2012), we also included a sample of undergraduate students from a large public university in the west coast of the United States for study #3 (with the image depicting a genetically modified cat/mouse). There were 486 MTurk workers and 401 undergraduate students who completed study #3, all randomly assigned to the 28 experimental conditions. Compared to MTurk workers, students were significantly younger ($M_{students}$ = 21.24, SD = 2.58, t = –27.00, $p < .001$), with more women (71.54%, Pearson $\chi^2 = 45.11$, $p < .001$), had lower Internet skills ($M_{students}$ = 3.31, SD = 0.84, t = –12.23, $p < .001$), less photography/digital imaging experiences ($M_{students}$ = 2.74, SD = 0.75, t = –3.70, $p < .001$), and less likely to have a Twitter account (45.64%, Pearson $\chi^2 = 17.91$, $p < .001$) but equally likely to have a Facebook account (92.02%, Pearson $\chi^2 = 1.64$, $p = .201$). They were also marginally more credulous than MTurk workers ($M_{students}$ = 1.95, SD = .95, t = 1.87, p = .06).[1] In the following section, only MTurk sample results are reported.

Results

There were slightly more men (N = 1902, 54.72%) than women (N = 1548, 44.53%) among those who completed the study. Participants were between 20 and 87 years old (one participant reported being 11 years old and was removed as all participants are required to be 18 or older to enter the study), with a mean age of 34.71 years (SD = 11.16). The largest household income category was less than 30,000 US dollars annually. Participants were well-educated, with 89.8% reporting at least some college or above. Detailed demographic statistics are reported in Table 2.

Overall, we observed significant differences in the average credibility judgment of the six images, as expected. The mean credibility ratings on a 7-point scale for each of the images are: 4.65 (SD = 1.19, bridge collapse), 3.86 (SD = 1.74, gay couple adopting children), 1.83 (SD = 0.96, genetically modified mouse), 3.08 (SD = 1.66, a school in Africa), 4.06 (SD = 1.35, Syrian bombing), and 2.29 (SD = 1.32, Hispanic politician). The descriptive statistics and correlations are reported in Table 3.

To test all hypotheses and answer research questions, we ran an analysis of covariance (ANCOVA) for all participants (N = 3476), with all four experimental factors (source trustworthiness, source and media type, intermediary, and bandwagon). The participant's sex and the image tested were considered as fixed factors. The covariates were the participant's age, digital imaging experience, Internet skills, and favorable attitude toward the issue. An interaction term between source trustworthiness and intermediary was also included.

H1 predicted that images from highly trustworthy sources are evaluated as more credible than those from less trustworthy sources. H1 was not supported, as source trustworthiness did not have a significant main effect in the whole model, $F(1, 3449) = 1.64$, $p = .20$.

Table 2. Descriptive statistics of participant demographics ($N = 3476$).

		Total	Percent
Sex	Men	1902	54.72
	Women	1548	44.53
	Not disclosed	26	0.75
Age		$M = 34.71$	$SD = 11.16$
Race	White/Caucasian	2596	74.68
	African American	241	6.93
	Hispanic	199	5.72
	Asian	325	9.35
	Native American	19	.55
	Pacific Islander	11	.32
	Other	62	1.78
	Rather not to disclose	23	.66
Income	Less than US$30,000	819	23.56
	US$30,000–US$39,999	480	13.81
	US$40,000–US$49,999	399	11.48
	US$50,000–US$59,999	391	11.25
	US$60,000–US$69,999	286	8.23
	US$70,000–US$79,999	268	7.71
	US$80,000–US$89,999	145	4.17
	US$90,000–US$99,999	164	4.72
	US$100,000 or more	462	13.29
	Rather not to disclose	62	1.78
Education	Less than High School	21	.60
	High School/GED	319	9.18
	Some College	893	25.69
	2 year College Degree	379	10.90
	4 year College Degree	1340	38.55
	Masters Degree	396	11.39
	Doctoral Degree	50	1.44
	Professional Degree (JD, MD)	64	1.84
	Rather not to disclose	14	.40

H2a predicted that images from news organizations are perceived as more credible than those from individuals. H2b predicted that images from an organization's official website will be perceived as more credible than those from their social media accounts. We tested the main effect of source and media type, as well as planned contrasts between the three levels within the factor (news organization website, news organization social media, individual social media). The results showed that the main effect was nonsignificant for the whole sample, $F(2, 3449) = 1.75$, $p = .17$, so were the planned contrasts. As a result, H2a and H2b were not supported.

H3 predicted that images from more credible intermediaries will be perceived as more credible than those from less credible intermediaries, while RQ1 asked if having an

Table 3. Correlations of continuous variables ($N = 3476$).

		Mean	SD	1	2	3	4	5	6
1	Age	35.71	11.16						
2	photo_experience	2.82	0.79	−0.09**					
3	pro issue attitude	4.14	1.81	-.07**	0.004				
4	internet_skills	4.04	0.82	-.04*	.33**	.04*			
5	photo_credibility	3.50	1.73	.07**	-.11**	.05**	-.05**		
6	TW intensity (account holders only, $N = 1927$)	2.58	1.06	-.08**	.10**	−0.01	0.01	0.03	
7	FB intensity (account holders only, $N = 3114$)	3.04	1.03	.12**	.04*	-.04*	-.04**	0.03	.19**

**$p < .01$ (2-tailed). *$p < .05$ (2-tailed).

intermediary affects image credibility. The factor intermediary did not have a significant main effect, $F(2, 3449) = 0.97$, $p = .38$. Subsequent planned contrasts among the three levels (no intermediary, low trustworthiness, and high trustworthiness) yielded nonsignificant results. Therefore, H3 was not supported, and the answer to RQ1 was negative. RQ2 explored the potential interaction between source trustworthiness and intermediary. We again found no significant interaction in all models.

H4 predicted that images with higher levels of bandwagon cues such as shares and favorites will be perceived as more credible than those with lower levels of bandwagon cues. The main effect of bandwagon cues was nonsignificant in the whole model, $F(1, 3449) = 0.04$, $p = .85$, as well as in both subsamples. H4 was therefore not supported.

H5a predicted that people with greater amounts of photography experience and digital imaging skills will perceive images as less credible compared to people with less skill or experience. This hypothesis was supported in the whole model, $F(1, 3449) = 12.38$, $p < .001$. H5b predicted that people with greater levels of Internet skills will perceive images as less credible compared to people with lower skills. This prediction was also supported, $F(1, 3449) = 6.79$, $p = .01$.

To investigate further how Facebook and Twitter use in particular may play a role in credibility judgment of online images using either Facebook or Twitter mockups, we divided the participants into two subsamples: the Twitter sample is based on people who were exposed to Figures 1, 3, and 6 (bridge collapse in China, genetically modified cat/mouse, Hispanic politician), where the Twitter interface was used in the mockup; The Facebook sample is based on people who were exposed to Figures 2, 4, or 5 (gay couple, African school, and Syrian bombing). We ran separate ANCOVAs on both samples, using the same design as the whole model, first adding whether participants have a Facebook/Twitter account (binary variable), and if they do, their Facebook/Twitter use intensity measure, respectively. This resulted in two ANCOVA models for the Facebook subsample and two ANCOVA models for the Twitter subsample (Table 4).

H6a predicted that people who use Facebook more will perceive images as less credible compared to people who use Facebook less. This hypothesis did not receive support as Facebook use intensity was not associated with credibility rating, Model 3:

Table 4. ANCOVA predicting image credibility.

Variables	Whole Sample		Facebook subsample				Twitter subsample			
	Model 1		Model 2		Model 3		Model 4		Model 5	
	F	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Corrected Model	104.44	<.001	11.99	<.001	10.94	<.001	146.96	<.001	86.62	<.001
Intercept	362.84	<.001	137.14	<.001	91.25	<.001	262.86	<.001	108.12	<.001
source & media type	1.75	.17	3.40	0.03	2.28	0.10	2.13	0.12	0.91	0.40
bandwagon	0.04	.85	0.00	0.98	0.06	0.81	0.03	0.86	1.18	0.28
source trustworthiness	1.64	.20	3.22	0.07	1.58	0.21	0.05	0.83	0.21	0.65
intermediary	0.97	.38	1.09	0.34	0.89	0.41	1.02	0.36	0.19	0.83
photo_experience	12.38	<.001	3.01	0.08	3.51	0.06	11.45	0.00	4.49	0.03
internet_skills	6.79	.01	4.23	0.04	1.47	0.23	4.68	0.03	1.37	0.24
female	2.63	.10	7.46	0.01	7.59	0.01	2.24	0.14	2.43	0.12
age	44.08	<.001	39.61	<.001	36.29	<.001	6.76	0.01	3.36	0.07
source_trustworthiness * intermediary	0.87	.42	2.44	0.09	1.79	0.17	0.52	0.60	0.32	0.73
pro_attitude	9.00	.00	10.94	0.001	8.74	0.003	0.28	0.60	0.01	0.92
picture_condition	350.68	<.001	51.45	<.001	48.02	<.001	1099.50	<.001	647.98	<.001
FB_account			0.14	0.71						
FB intensity					0.86	0.35				
TW_account							1.15	0.28		
TW_intensity									5.98	0.02
R^2	.35	.35	.10		.10		.58		.59	
df	3449		1701		1535		1748		999	

$F(1, 1535) = 0.86$, $p = 0.35$. H6b predicted that people who use Twitter more will perceive images as less credible compared to people who use Twitter less. This hypothesis was supported, as Twitter use intensity was significant, Model 5: $F(1, 999) = 5.98$, $p = .02$.

H7 predicted that people's support of the issue depicted in the image is positively related to their credibility rating of the image. This hypothesis received strong support in the whole sample, $F(1, 3449) = 9.00$, $p < .001$, and Facebook subsample, Model 2: $F(1, 1701) = 10.94$, $p = .001$; Model 3, $F(1, 1535) = 8.74$, $p = .003$.

Finally, participants' sex and age were included as controls. Age showed a strong main effect across the board, Model 1: $F(1, 3449) = 44.08$, $p < .001$. Sex was significant in the Facebook subsample, Model 2: $F(1, 1701) = 39.61$, $p < .001$; Model 3, $F(1, 1535) = 36.29$, $p < .001$, but not significant in the whole sample, Model 1: $F(1, 3449) = 2.63$, $p = .105$, or the Twitter subsample.

Discussion

As tools for creating and manipulating digital images become increasingly commonplace and easy to use, fake images continue to propagate across social media platforms and contemporary media environment, influencing the viewers and posing a significant sociopolitical threat around the world. It is thus imperative to better understand how people evaluate the credibility of online images. This study reports the findings from a large-scale experiment on image credibility evaluations on the Web, conducted on Amazon MTurk. Based on previous work on social and cognitive heuristics for evaluating online credibility, we tested the effects of several features such source, intermediary, and the background and skills of the viewers on assessing the credibility of images online. The results were consistent across all six images tested, showing that viewers' Internet skills, digital imaging experiences, social media use, and pro-issue attitude are significant predictors of image credibility evaluation. However, none of the image context features tested—for example, where the image was posted or and how many people liked it—had an impact on participants' credibility judgments. Our findings also reveal that credibility evaluations are far less impacted by the content of an online image. Instead they are influenced by the viewers' backgrounds, prior experiences, and digital media literacy.

This study contributes critical insights to image credibility research. Past studies reported that people generally believe that they are rarely capable of identifying fake images as such (Farid and Bravo, 2010), and that images are generally considered trustworthy (Kasra et al., 2018; Nightingale et al., 2017). Yet the credibility variance of images was limited in these studies, either by the measurement scale employed (binary yes/no), or by the topics and contexts of the images. To ensure a large variance, our study purposefully chose six fake images depicting a wide range of issues. Each image was forged using various image-manipulation techniques (e.g. composition, elimination, retouching) and exhibited different levels of sophistication. Contrary to what the previous studies suggested, our results show that people are not as gullible in evaluating image credibility on the Web. Our participants rated four images as fake or manipulated (below 4 on a 7-point scale). The other two images were rated only a little above the midpoint.

This result indicates that participants, no matter how careless or distracted they may be, can still be discerning consumers of digital images.

Compared to previous research, our study implemented three notable changes. First, our design recognized that image consumption and evaluation on the Internet are always contextual rather than occurring in a vacuum. We therefore provided brief textual information about each fabricated image, similar to how images are usually presented and viewed online. Second, taking into account that online information is continuously shared and reshared by different sources, we explicitly manipulated and tested whether the existence and trustworthiness of an intermediary had any bearing on image credibility evaluation. Third, we adopted a measure of credibility (6 items, on a 7-point scale) that is more nuanced than a binary yes/no choice, which was used in the study by Nightingale et al. (2017). Our scale has better reliability and validity than a binary measure, which is prone to false positive and false negative results.

The most significant discovery of our study is that viewers' skills and experience greatly impact their image credibility evaluations. The more knowledge and experience people have with the Internet, digital imaging and photography, and the online media platforms, the better they are at evaluating image credibility. Our results suggest that to mitigate the potential harm caused by fake images online, the best strategy is investing in educational efforts to increase users' digital media literacy. Meanwhile, issue attitude has a significant effect as well. This is consistent with the confirmation bias found in many similar studies (e.g. Knobloch-Westerwick et al., 2015) that people are more likely to accept an image as real if it aligns with their prior beliefs. This finding could explain why fake news spreads so readily in social media settings.

Several well-researched social and cognitive heuristic cues found in online credibility research (e.g. source trustworthiness, media platform, and bandwagon cues) did not have any significant effect on image credibility. Although surprising, this result does not mean that the process of image credibility evaluation is inherently different from the process of judging online information, neither does it mean that people use very different heuristics. We speculate that workers on MTurk might have been rushing through the experiment without paying enough attention to the various source, intermediary, and bandwagon cues. MTurk participants were certainly not motivated to pay attention (Antin and Shaw, 2012), as they were compensated by the completion of the task, regardless of the response quality. We included a few quality-control mechanisms, such as a 30 second minimum stay time on the image page before participants could advance to the next page.[2] However, it can be argued that a rushed, careless scan without motivation to consider various cues is indeed how people consume news and images in today's media environment. In this regard, the MTurk workers' behaviors may be representative of people's actual behaviors online.

Limitation and future research

This study has a number of limitations. We recruited participants from MTurk, a widely used platform for social science studies like ours. Although the platform allowed us to recruit a reliable, inexpensive, and demographically diverse sample, workers on MTurk

are still not representative of the general population. We found them to have good self-reported Internet skills (M=4.04 on a 5-point scale), compared to a student sample in our study and those reported previously (Hargittai and Hsieh, 2012; Hargittai and Shaw, 2015), perhaps unsurprisingly as they participated in an online labor marketplace. Still, research found that the MTurk samples are slightly more diverse demographically than standard Internet samples, and a lot more diverse than American college samples (Buhrmester et al., 2011). The second limitation is the lack of manipulation checks in our design. As a result, we do not know for sure whether observed results stemmed from lack of attention (e.g. participants did not notice the purported source of an image was New York Times) or the factor itself (e.g. participants did not think New York Times was a credible source). Given the pretest results, we believe the former is more likely than the latter, yet this remains a minor threat to validity. Third, our study was cross-sectional so causality cannot be ascertained, although we believe most variables capture pre-existing states and habits (Internet skills, digital imaging experiences) that are unlikely to change due to image evaluation tasks.

In addition, we purposefully included only fake images in the credibility evaluation task and excluded unaltered and/or misattributed original images. Although this approach eliminated the risk of participants being already familiar with the stimuli, it only tested participants' suspicion when confronted with an image that was actually fake. In other words, for those participants who rated fake images as less credible, we could not determine whether they were truly capable at evaluating image credibility, or just being more skeptical in general. However, given the amount of misinformation and disinformation in today's media environment, being skeptical is arguably the crucial first step in all credibility evaluation tasks. As Rheingold (2012) argues, "the heuristic for crap detection is to make skepticism your default" (p. 77). Meanwhile, the fake images used in this study were all forgeries of sufficient quality so as not to be immediately distinguishable from authentic images. Previous work showed that participants failed to identify the images as fake and even when told that the images were fake they failed to correctly identify what image areas had been manipulated (Kasra et al., 2018). Inability to distinguish between compelling fakes and authentic images implies that results would be the same regardless of whether the study was done using authentic images, compelling fakes, or a mixture. Nevertheless, future research should test users' ability to evaluate unaltered and misattributed images as well as fakes of varying quality.

Even though our experiment aimed to be comprehensive, it still left out a few important factors. For example, we did not manipulate the recency of images posted (all images were presented with random 2015 dates), which could influence credibility judgments (Westerman et al., 2014). Furthermore, considering that people are more likely to consume news on social media sites instead of traditional channels, and that their networks of "friends" will play an important role in information diffusion, a productive future research direction is to examine credibility judgment in participants' naturalistic social network environment. This will allow the study to factor in the endorsements and aggregate ratings from the participants' self-curated network of friends and contacts. We also focused only on image consumption on the desktop while people increasingly access news on mobile devices (Fedeli and Matsa, 2018). How the parameters of mobile devices may impact credibility judgment remains a fruitful future direction. Operationally, our

study design could be further improved by swapping the order of the Internet and digital skills questions with the fake image to eliminate potential priming effects. Forcing the participants to stay on the image page for 30 seconds was also a less than ideal approach to ensuring sufficient time for evaluation. Furthermore, participants' Internet and digital photography skills were self-reported rather than measured objectively, although evidence shows that the likelihood of participants misreporting their Internet skills is low (Hargittai, 2009). Future research is encouraged to address the above operational concerns to further improve validity.

Conclusion

In the age of fake news and alternative facts, the risks and dangers associated with ill-intentioned individuals or groups easily routing forged visual information through computer and social networks to deceive, cause emotional distress, or to purposefully influence opinions, attitudes, and actions have never been more severe. This article details an online experiment to probe how people respond to and evaluate the credibility of images in online environments. Through a series of between-subjects factorial experiments that randomly assigned participants on Mechanical Turk to rate the credibility of fake image mockups, we found that image characteristics such as where it is published and how many people shared it do not matter. Instead, participants' Internet skills, digital imaging experiences, social media use, and pro-issue attitude are significant predictors of credibility evaluation of online images.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by National Science Foundation grants CNS-1444840 and CNS-1444861.

Notes

1. We ran the same ANCOVA analyses on the student sample and found little difference from the MTurk sample.
2. We also ran models including the total time spent on the image page as a covariate, which did not change results.

ORCID iD


Cuihua Shen https://orcid.org/0000-0003-1645-8211

References

Allcott H and Gentzkow M (2017) Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives* 31: 211–236.

Antin J and Shaw A (2012) Social desirability bias and self-reports of motivation: a study of Amazon Mechanical Turk in the US and India. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. Austin, TX, 5–10 May, pp. 2925–2934. New York: ACM Press.

Bailenson JN, Iyengar S, Yee N, et al. (2008) Facial similarity between voters and candidates causes influence. *Public Opinion Quarterly* 72: 935–961.

Buhrmester M, Kwang T and Gosling SD (2011) Amazon's Mechanical Turk: a New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science* 6: 3–5.

Choi W and Stvilia B (2015) Web credibility assessment: conceptualization, operationalization, variability, and models. *Journal of the Association for Information Science and Technology* 66: 2399–2414.

Ellison NB, Steinfield C and Lampe C (2007) The benefits of Facebook "friends": social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication* 12: 1143–1168.

Farid H and Bravo MJ (2010) Image forensic analyses that elude the human visual system. *Proceedings of SPIE* 7541: 754106.

Fedeli S and Matsa KE (2018) Use of mobile devices for news continues to grow, outpacing desktops and laptops. Available at: http://www.pewresearch.org/fact-tank/2018/07/17/use-of-mobile-devices-for-news-continues-to-grow-outpacing-desktops-and-laptops/

Flanagin AJ and Metzger MJ (2007) The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society* 9: 319–342.

Fogg BJ, Soohoo C, Danielson DR, et al. (2003) How do users evaluate the credibility of Web sites? A study with over 2,500 participants. In: *Proceedings of the 2003 conference on designing for user experiences*. San Francisco, CA, 6–7 June, pp. 1–15. New York: ACM Press.

Friggeri A, Adamic LA, Eckles D, et al. (2014) Rumor cascades. In: *Proceedings of the 8th International AAAI conference on weblogs and social media (ICWSM)*, pp. 101–110. Menlo Park, CA: AAAI.

Gerlitz C and Helmond A (2013) The like economy: social buttons and the data-intensive web. *New Media & Society* 15: 1348–1365.

Greer JD and Gosen JD (2002) How much is too much? *Visual Communication Quarterly* 9: 4–13.

Gupta A, Lamba H, Kumaraguru P, et al. (2013) Faking Sandy: characterizing and identifying fake images on Twitter during Hurricane Sandy. In: *Proceedings of the 22nd international conference on World Wide Web companion*, Rio de Janeiro, Brazil, 13 May, pp. 729–736. Rio de Janeiro, Brazil: International World Wide Web Conferences Steering Committee.

Hargittai E (2009) An update on survey measures of web-oriented digital literacy. *Social Science Computer Review* 27: 130–137.

Hargittai E and Hsieh YP (2012) Succinct survey measures of web-use skills. *Social Science Computer Review* 30: 95–107.

Hargittai E and Shaw A (2015) Mind the skills gap: the role of Internet know-how and gender in differentiated contributions to Wikipedia. *Information, Communication & Society* 18: 424–442.

Kasra M, Shen C and O'Brien JF (2018) Seeing is believing: how people fail to identify fake images on the web. In: *Extended abstracts of the 2018 CHI conference on human factors in computing systems*, Montreal QC, Canada, 21–26 April, Paper No. LBW516. New York: ACM Press.

Knobloch-Westerwick S, Johnson BK and Westerwick A (2015) Confirmation bias in online searches: impacts of selective exposure before an election on political attitude strength and shifts. *Journal of Computer-Mediated Communication* 20: 171–187.

Knobloch-Westerwick S, Sharma N, Hansen DL, et al. (2005) Impact of popularity indications on readers' selective exposure to online news. *Journal of Broadcasting & Electronic Media* 49: 296–313.

Light K (2004) *Fonda, Kerry and photo fakery*. The Washington Post February 28: A21.

Lee JY and Sundar SS (2013) To Tweet or to Retweet? That Is the Question for Health Professionals on Twitter. *Health Communication* 28: 509–524.

Metzger MJ (2007) Making sense of credibility on the web: models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology* 58: 2078–2091.

Metzger MJ, Flanagin AJ and Medders RB (2010) Social and heuristic approaches to credibility evaluation online. *Journal of Communication* 60: 413–439.

Metzger MJ, Flanagin AJ, Eyal K, et al. (2003) Credibility for the 21st century: integrating perspectives on source, message, and media credibility in the contemporary media environment. *Annals of the International Communication Association* 27: 293–335.

Mitchell A, Gottfried J, Kiley J, et al. (2014) *Political polarization & media habits*. Available at: http://www.journalism.org/2014/10/21/political-polarization-media-habits/#trust-and-distrust-liberals-trust-many-conservatives-trust-few

Morris MR, Counts S, Roseway A, et al. (2012) Tweeting is believing? understanding microblog credibility perceptions. In: *Proceedings of the ACM 2012 conference on computer supported cooperative work*. Seattle, WA, 11–15 February, pp. 441–450. New York: ACM Press.

Nash RA, Wade KA and Brewer RJ (2009) Why do doctored images distort memory? *Consciousness and Cognition* 18: 773–780.

Nightingale SJ, Wade KA and Watson DG (2017) Can people identify original and manipulated photos of real-world scenes? *Cognitive Research: Principles and Implications* 2: 30.

Rheingold H (2012) *Net Smart: How to Thrive Online*. Cambridge, MA: The MIT Press.

Sacchi DLM, Agnoli F and Loftus EF (2007) Changing history: doctored photographs affect memory for past public events. *Applied Cognitive Psychology* 21: 1005–1022.

Smith C (n.d.) Reader's digest trust poll: the 100 most trusted people in America. Available at: https://www.rd.com/culture/readers-digest-trust-poll-the-100-most-trusted-people-in-america/

Stewart N, Ungemach C, Harris AJ, et al. (2015) The average laboratory samples a population of 7,300 Amazon Mechanical Turk workers. *Judgment and Decision Making* 10: 479–491.

Sundar SS (2008) The MAIN model: a heuristic approach to understanding technology effects on credibility. In: Metzger MJ and Flanagin AJ (eds) *Digital Media, Youth, and Credibility*. Cambridge, MA: The MIT Press, pp. 73–100.

The Guardian (2014) Award-winning photographer dumped for altering single Syria image. *The Guardian*. Available at: https://www.theguardian.com/media/2014/jan/23/photographer-dumped-altering-syria-image

Treier S and Hillygus DS (2009) The nature of political ideology in the contemporary electorate. *The Public Opinion Quarterly* 73: 679–703.

Tseng S and Fogg BJ (1999) Credibility and computing technology. *Communication of the ACM* 42: 39–44.

Wade KA, Garry M, Read JD, et al. (2002) A picture is worth a thousand lies: using false photographs to create false childhood memories. *Psychonomic Bulletin & Review* 9: 597–603.

Westerman D, Spence PR and Van Der Heide B (2012) A social network as information: the effect of system generated reports of connectedness on credibility on Twitter. *Computers in Human Behavior* 28: 199–206.

Westerman D, Spence PR and Van Der Heide B (2014) Social media as information source: recency of updates and credibility of information. *Journal of Computer-Mediated Communication* 19: 171–183.

Wineburg S and McGrew S (2016) Evaluating information: the cornerstone of civic online reasoning. Available at: https://sheg.stanford.edu/upload/V3LessonPlans/Executive%20Summary%2011.21.16.pdf

Author biographies

Cuihua Shen is an associate professor at the Department of Communication and co-founder of the Computational Communication Research Lab, UC Davis. Her research interests include digital media literacy, online and mobile social networks, and computational social science.

Mona Kasra is an assistant professor of Digital Media Design at the University of Virginia (UVa). Her research is centered around the power and politics of emerging media and its impact on art, culture, and society. In 2016, she served as Conference Chair at ACM SIGGRAPH, the world's largest, most influential annual conference on the theory and practice of computer graphics and interactive techniques.

Wenjing Pan completed this project as a PhD student in the Department of Communication at UC Davis. She is currently an assistant professor at the School of Journalism and Communication, Renmin University of China. Her research interests include online supportive communication, health communication, and computational social science.

Grace A Bassett completed this project as a PhD student in the Department of Communication at UC Davis. Her research interests focus on social networks, teams, language, and computational social science. She is currently working as a User Experience Researcher at Facebook.

Yining Malloch is a PhD candidate at the Department of Communication, UC Davis. Her research interests include computer-mediated social interaction and health communication.

James F O'Brien is a professor of Computer Science at the University of California, Berkeley. His research interests include media forensics, computer graphics, physically based animation, and simulation of physical systems. Professor O'Brien is a fellow of the Sloan, Okawa, and Hellman foundations and an ACM Distinguished Scientist, has been selected as one of Technology Review's TR-100, and received an Academy Award from the Academy of Motion Picture Arts and Sciences.