Isolating the Effects of Web Page Visual Appearance on the Perceived Credibility of Online News among College Students

Jacob O. Wobbrock,[1] Anya K. Hsu,[1*] Marijn A. Burger,[2*] Michael J. Magee[1]

[1] The Information School | DUB Group
University of Washington
Seattle, WA, USA 98195
{ wobbrock, anyahsu, mjm53}@uw.edu

[2] Department of Computing & Software Systems
University of Washington, Bothell
Bothell, WA, USA 98011
marijn@uw.edu

ABSTRACT

Online news sources have transformed civic discourse, and much has been made of their credibility. Although web page credibility has been investigated generally, most work has focused on the credibility of web page *content*. In this work, we study the isolated *appearance* of news-like web pages. Specifically, we report on a laboratory experiment involving 31 college students rating the perceived credibility of news-like web pages devoid of meaningful content. These pages contain only “lorem ipsum” text, indistinct videos and images, non-functional links, and various font settings. Our findings show that perceived credibility is indeed affected by some purely presentational factors. Specifically, video presence increased credibility, while large fonts and having no images reduced credibility. Having a few, but not too many, images increased credibility for short articles, especially in the presence of large fonts. We also conducted follow-up interviews, which revealed that participants noticed images, videos, and font sizes when making credibility judgments, corroborating our quantitative experimental results.

CCS CONCEPTS

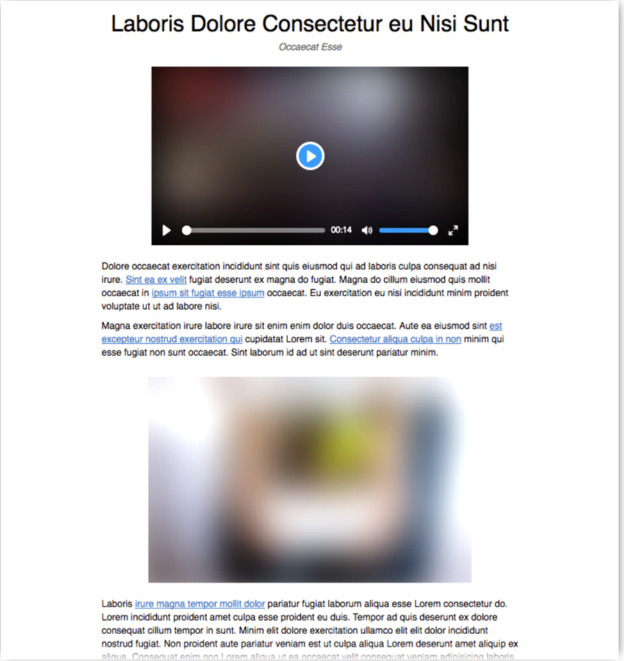
• **Information systems~Web interfaces** • Human-centered computing~Web-based interaction • Human-centered computing~Empirical studies in HCI.

KEYWORDS

Believability; credibility; trust; content; visual appearance; visual presentation; web page; video; images; fonts; online news.

ACM Reference format:
Jacob O. Wobbrock, Anya K. Hsu, Marijn A. Burger, and Michael J. Magee. 2019. Isolating the Effects of Web Page Visual Appearance on the Perceived Credibility of Online News among College Students. In *Proceedings of ACM Hypertext (HT’19), September 17-20, 2019, Hof, Germany*, ACM, New York, New York, USA. 10 pages. https://doi.org/10.1145/3342220.3343663

HT ’19, September 17–20, 2019, Hof, Germany
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-6885-8/19/09…$15.00
https://doi.org/10.1145/3342220.3343663



Laboris Dolore Consectetur eu Nisi Sunt

Occaecat Esse

Dolore occaecat exercitation incididunt sint quis eiusmod qui ad laboris culpa consequat ad nisi irure. Sint ea ex velit fugiat deserunt ex magna do fugiat. Magna do cillum eiusmod quis mollit occaecat in ipsum sit fugiat esse ipsum occaecat. Eu exercitation eu nisi incididunt minim proident voluptate ut ut ad labore nisi.

Magna exercitation irure labore irure sit enim enim dolor duis occaecat. Aute ea eiusmod sint est excepteur nostrud exercitation qui cupidatat Lorem sit. Consectetur aliqua culpa in non minim qui esse fugiat non sunt occaecat. Sint laborum id ad ut sint deserunt pariatur minim.

Laboris irure magna tempor mollit dolor pariatur fugiat laborum aliqua esse Lorem consectetur do. Lorem incididunt proident amet culpa esse proident eu duis. Tempor ad quis deserunt ex dolore consequat cillum tempor in sunt. Minim elit dolore exercitation ullamco elit elit dolor incididunt nostrud fugiat. Non proident aute pariatur veniam est ut culpa aliqua Lorem deserunt amet aliquip ex aliqua. Consequat enim non Lorem aliqua ut ea occaecat velit consequat veniam adipisicing laboris

Figure 1. An example web page generated in our study. The title and body fonts are sans serif 38 pt. and 16 pt., respectively. The word count for the whole article is 644, with some hyperlinked words visible in this screen shot. A heavily blurred video and image are also present.

1 INTRODUCTION

Web-based news articles, perhaps more than other forms of news, have immense flexibility in their design. Unconstrained by the physical limitations of print newspapers, a web-based news article can take on virtually any appearance its creator wishes. Also, with today’s authoring tools, far more people can create and publish “news” sites than produce or distribute printed newspapers. The range and influence of web-based news was seen clearly in the “fake news” epidemic promulgated by social media during the 2016 U.S. election [1]. American citizens, and others, widely shared misleading or false articles because they looked convincing and affirmed their viewpoints or biases.

College students are among the heaviest users of social media and online news in the United States, and social media is a key link to news for many of them. According to a recent Pew

[*] These authors contributed equally to this work.

Research Center study [3], college-educated Internet users are approximately 50% more likely to use social media than their peers, and users between the ages of 18 and 29 are almost twice as likely as any other demographic to use social media. Given the significant role of college students in promulgating news via social media, there is a need to better understand this demographic and their perceptions of news on the Web.

Many factors contribute to the proliferation of web-based news articles, including their content, timing, and appearance. Although much has been made of web page *content* [6,31,43], much less is known about how the mere *appearance* of an online news source—isolated from any particular content—might contribute to its perceived credibility. The goal of this work is to understand which *purely presentational factors* affect the perceived credibility of news-like web pages, and how.

To achieve this goal, we built a web-based system called *Pyrite* capable of generating news-like web pages that were devoid of any meaningful content (see, *e.g.*, Figure 1). By "devoid of any meaningful content," we mean that all text generated was from the first-century B.C. "lorem ipsum" source by Cicero [24]; that videos and images were swirls of blurry colors; and that hyperlinks were all "dead" (*i.e.*, non-functional). We conducted an experiment in which we controlled or randomized presentational factors, including font sizes and serifs, the number of images and videos, video placements, the number of words in the article, and the density of hyperlinks. We had 31 college-age participants rate how credible they felt each of 24 news-like web pages seemed to them if they imagined the articles had actual content. We also measured how much time participants spent on each page making their assessments.

Our major finding, which was unanticipated, was that some presentational factors, even when isolated from content, *do* affect the perception of web page credibility. Specifically, video presence increased credibility, while large fonts and having no images reduced it. Having a few, but not too many, images increased credibility for short articles, especially in the presence of large fonts. We also conducted interviews with each participant after their experiment session. Our main finding was that participants said they noticed images, videos, and font sizes most when making credibility judgments. These self-reports exactly corroborated our quantitative experimental results.

The contribution of this work is our empirical findings from our study of isolated presentational factors in news-like web pages. These findings can inform the design of online news, and can also inform citizens as to the effects of presentation on their perception of news sources. To the best of our knowledge, our work is the first to isolate web page appearance from content in the study of online news credibility.

2 RELATED WORK

In this section, we review work on source and message credibility, website credibility, and the impact of online news. For our purposes, and in keeping with prior work on website credibility [8], we adopt a definition of credibility as "believable or trustworthy." Interestingly, the root word for credibility is from the Latin *credere*, meaning "to believe." We do not make a distinction between credibility and believability, allowing, as others have done before us (*e.g.*, [8]), for "credible information" to be regarded equivalently as "believable information."

2.1 Source and Message Credibility

Source credibility has been studied in various contexts. Hass [15] defined a "credible source" as one that conveys accurate information, and does so without bias. When evaluated by television news audiences, source credibility has been shown to be based on apparent expertise and trustworthiness [19].

When the source of a message is unknown, evaluations of credibility tend to be based on the message itself [33]. Slater and Rouner [37] have shown that message credibility interacts with source credibility to produce overall credibility perceptions. Metzger *et al.* [28] argue that message structure, message content, language intensity, and message clarity are the key factors constituting message credibility. Judgments of message credibility are also shaped by the types of information conveyed [6]. Olaisen [30] showed that factors related to the source or content of a message are distinct from factors related to the message's medium and its design features, the latter related to what is being investigated in our work here.

2.2 Website Credibility

Given the ever-increasing amount of information online, it is no surprise that website credibility has been an active topic of research for some time. Writers and researchers have long sought to understand how website credibility judgments are formed. Easy access to web hosting services means that online information is not governed by the same "professional gatekeepers" as print media, and therefore has a much higher risk of being inaccurate [23,27]. Research has shown that Internet users generally lack both the motivation and skills to verify the information they find online [2,26]. Flanagin and Metzger [6] showed that even when users do possess the skills to verify information on the Internet, in practice, they rarely bother doing so.

Studies of website credibility have shown that credibility judgments are rapid and complex, incorporating multiple dimensions simultaneously [10]. For example, Freeman and Spyridakis [12] showed that readers evaluate credibility based on objective judgements about the information's accuracy as well as subjective judgements about the information's "trustworthiness, expertise, and attractiveness." Flanagin and Metzger [7] and Furman [13] argue that credibility perceptions are based more on visual attributes of a web page, like design features and apparent site complexity, rather than knowledge of the source of the information. Tuch *et al.* [41] also find that visual complexity plays a role in forming aesthetic judgments of websites, and Tractinsky *et al.* [40] argue that first impressions of website attractiveness affect perceptions of trustworthiness. (Beldad *et al.* [4] provide a review of factors related to website trustworthiness generally.)

Although we are unaware of any studies that focused only on presentational factors as we do here, some prior work has included visual elements among other factors when investigating website credibility. Fogg *et al.* [9] conducted a study in which 46.1% of their respondents mentioned looking at the high-level design of a website when forming credibility judgments. As Fogg *et al.* observed, "No matter how good a site's content, the visual aspects of a web site will have a significant impact on how people assess credibility." Robins and Holmes [32] expanded on this finding, establishing that content with a "higher aesthetic treatment" was perceived as more credible. A literature review by Wathen and Burkell [43] established that to positively impact perceived credibility, a website must "emphasize a good interface and project a professional image, making use of established design principles." Stonewall and Dorneich [39] studied web page appearance, gathering users' ratings of visual attributes (*e.g.*, colors, shapes, images) as they relate to professionalism and gender. Kim and Moon [20] studied visual attributes in online banking, finding that professional looking graphics and colors increased perceived trustworthiness. Spillane *et al.* [38] conducted a crowdsourced credibility study of distorted websites where certain types of content were present or absent (*e.g.*, banner ads, share buttons, comment fields). Our current study affirms prior high-level findings of the importance of visual appearance on credibility, but our study goes further by isolating presentation from content, and by identifying *which* presentational aspects affect perceived credibility, and how much.

2.3 Importance of Online News

Many traditional print media have moved online, and as of August 2017, 43% of Americans report getting their news *primarily* from online sources [14]. As Burbules [5] observed even 20 years ago, the traditional reliance on established, reputable news sources for information has been diluted by the Web. The variety of news-like information available online has introduced a kind of "leveling effect" that gives all information, reliable or not, an equal level of accessibility, and thereby imbues all authors with the same initial semblance of credibility.

This trend culminated in the 2016 U.S. presidential election, during which the term "fake news" was coined due to the proliferation of misleading and inaccurate news articles online [18]. These fake news sources crafted volatile and biased stories about presidential candidates and other political figures, and such stories were then shared on social networks even more widely than the most popular mainstream news stories [35]. Surveys conducted at the time indicated that most people who read fake news articles believed them, never verifying the facts "reported" [36]. A recent study by Allcott and Gentzkow [1] established that the majority of American adults viewed and remembered at least one fake news story during the election, indicating the impact of fake news on the American electorate.

Given the impact of online news, the proliferation of this news via social media, the heavy social media use by college-aged adults, and the importance of the visual presentation of web pages in affecting people's credibility judgments, we sought to *isolate* just how much presentational factors affect credibility judgments. We next describe our study attempting to address this question.

3 STUDY METHOD

The purpose of our experiment was to isolate which presentational factors affect college students' perceived credibility of news-like web pages. We also investigated how presentational factors affected time-on-page, and conducted interviews after study sessions to understand participants' subjective perceptions.

3.1 Participants

We recruited 31 college students for our study. Thirteen participants were female (42%) and 18 were male (58%);[1] the mean age was 20.8 years old (*SD*=1.4). Three participants were recruited by speaking to a class at our local university and inviting students to participate in the study; the other 28 were approached in the university library. Each participant was compensated $10 USD.

As stated above, we wanted to focus on college students because of their high engagement with social media and online news [3]. Our participants were therefore limited to college-aged adults currently or recently enrolled in college. Therefore, our results might not generalize to people of other ages, or to non-college students of a similar age.

Of our 31 participants, 12 were arts or science majors, six were from information science, six were from engineering, and three were from business. The remaining four were from professional disciplines: education, pharmacy, and public health.

3.2 Apparatus

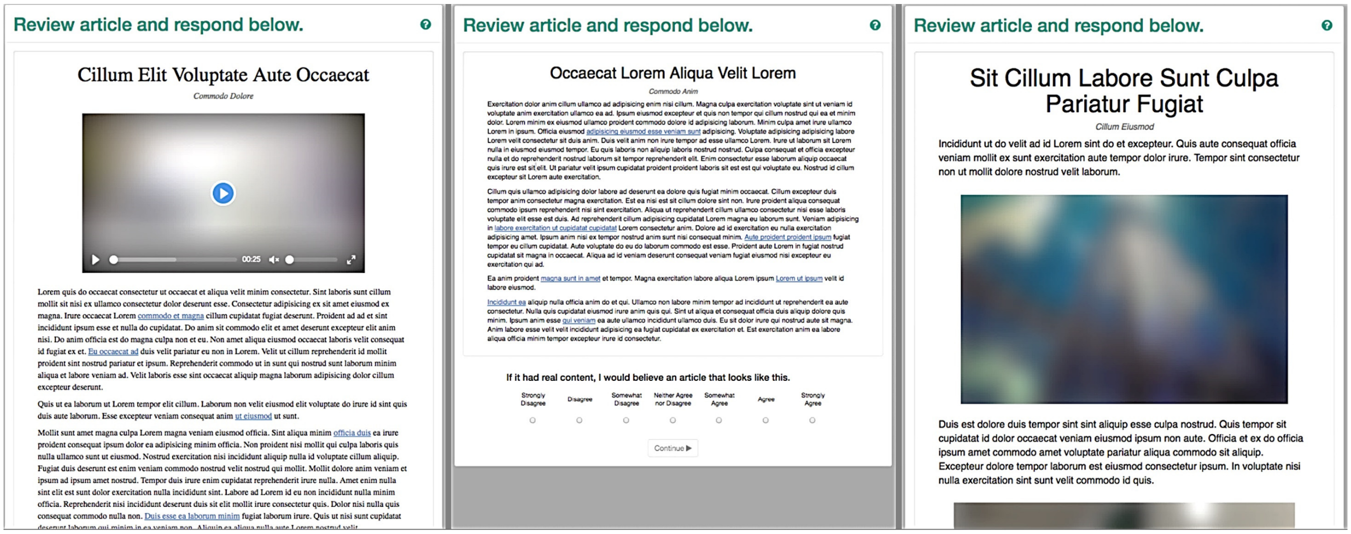
To run our experiment, we created a custom-built online testbed called Pyrite. Pyrite ran in a web browser and presented "content-free" news-like articles to participants (see, *e.g.*, Figure 1), recording their perceived credibility judgments and times-on-page. Pyrite displayed articles in the Google Chrome web browser in full-screen mode with no other windows open or applications running. Five of 31 participants used Macintosh desktops, desktop mice, and had 27" displays; the other 26 participants used Macintosh laptops, laptop trackpads, and had 13" displays.

Pyrite generated news-like articles based on real-world web page designs employed by actual online news sites. To arrive at the designs, we took the top 20 U.S. news websites from the Alexa rankings.[2] From there, we visited five random articles from each site's "news" or "world" sections. From this total sample of 100 news articles, we measured font faces and sizes, word counts, link densities, and image and video sizes, counts, and placements. We then built these values into Pyrite such that it used them when generating its news-like articles. (For specific values, see Section 3.4, below.)

As noted above, articles generated by Pyrite were intentionally devoid of meaningful content so that participants would have no content-based influences on their credibility judgments. If we *had* allowed discernable content into the articles, no matter what the topic chosen, that content would have introduced confounds that would have affected our ability to isolate purely presentational factors. Even innocuous-seeming topics like "cats" would appeal

[1] Options for gender included "male," "female," "non-binary," and "prefer not to respond." All participants chose either "male" or "female."

[2] https://www.alexa.com/topsites/category/Top/News



Review article and respond below.

Cillum Elit Voluptate Aute Occaecat

Commodo Dolore

Lorem quis do occaecat consectetur ut occaecat et aliqua velit minim consectetur. Sint laboris sunt cillum mollit sit nisi ex ullamco consectetur dolor deserunt esse. Consectetur adipisicing ex sit amet eiusmod ex magna. Irure occaecat Lorem commodo et magna cillum cupidatat fugiat deserunt. Proident ad ad et sint incididunt ipsum esse et nulla do cupidatat. Do anim sit commodo elit et amet deserunt excepteur elit anim nisi. Do anim officia est do magna culpa non et eu. Non amet aliqua eiusmod occaecat laboris velit consequat id fugiat ex et. Eu occaecat ad duis velit pariatur eu non in Lorem. Velit ut cillum reprehenderit id mollit proident sint nostrud pariatur et ipsum. Reprehenderit commodo ut in sunt qui nostrud sunt laborum minim aliqua et labore veniam ad. Velit laboris esse sint occaecat aliquip magna laborum adipisicing dolor cillum excepteur deserunt.

Quis ut ea laborum ut Lorem tempor elit cillum. Laborum non velit eiusmod elit voluptate do irure id sint quis duis aute laborum. Esse excepteur veniam consequat anim ut eiusmod ut sunt.

Mollit sunt amet magna culpa Lorem magna veniam eiusmod officia. Sint aliqua minim officia duis ea irure proident consequat ipsum dolor ea adipisicing minim officia. Non proident nisi mollit qui culpa laboris quis nulla ullamco sunt ut eiusmod. Nostrud exercitation nisi incididunt aliquip nulla id voluptate cillum aliquip. Fugiat duis deserunt est enim veniam commodo nostrud velit nostrud qui mollit. Mollit dolore anim veniam et ipsum ad ipsum amet nostrud. Tempor duis irure enim cupidatat reprehenderit irure nulla. Amet enim nulla sint elit est sunt dolor exercitation nulla incididunt sint. Labore ad Lorem id eu non incididunt nulla minim officia. Reprehenderit nisi incididunt deserunt duis sit elit mollit irure consectetur quis. Dolor nisi nulla quis consequat commodo nulla non. Duis esse ea laborum minim fugiat laborum irure. Quis ut nisi sunt cupidatat deserunt laborum qui minim in ea veniam non. Aliquip ea aliqua nulla aute Lorem nostrud velit

Review article and respond below.

Occaecat Lorem Aliqua Velit Lorem

Commodo Anim

Exercitation dolor anim cillum ullamco ad adipisicing enim nisi cillum. Magna culpa exercitation voluptate sint ut veniam id voluptate anim exercitation ullamco ea ad. Ipsum eiusmod excepteur et quis non tempor qui cillum nostrud qui ea et minim dolor. Lorem minim ex eiusmod ullamco proident commodo dolore id adipisicing laborum. Minim culpa amet irure ullamco Lorem in ipsum. Officia eiusmod adipisicing eiusmod esse veniam sunt adipisicing. Voluptate adipisicing adipisicing labore Lorem velit consectetur sit duis anim. Duis velit anim non irure tempor ad esse ullamco Lorem. Irure ut laborum sit Lorem nulla in eiusmod eiusmod tempor. Eu quis laboris non aliquip laboris nostrud nostrud. Culpa consequat et officia excepteur nulla et do reprehenderit nostrud laborum sit tempor reprehenderit elit. Enim consectetur esse laborum aliquip occaecat quis irure est sit elit. Ut pariatur velit ipsum cupidatat proident proident laboris sit est est qui voluptate eu. Nostrud id cillum excepteur sit Lorem aute exercitation.

Cillum quis ullamco adipisicing dolor labore ad deserunt ea dolore quis fugiat minim occaecat. Cillum excepteur duis tempor anim consectetur magna exercitation. Est ea nisi est sit cillum dolore sint non. Irure proident aliqua consequat commodo ipsum reprehenderit nisi sint exercitation. Aliqua ut reprehenderit cillum ullamco consectetur nisi esse laboris voluptate elit esse est duis. Ad reprehenderit cillum adipisicing cupidatat Lorem magna eu laborum sunt. Veniam adipisicing in labore exercitation ut cupidatat cupidatat Lorem consectetur anim. Dolore ad id exercitation eu nulla exercitation adipisicing amet. Ipsum anim nisi ex tempor nostrud anim sunt nisi consequat minim. Aute proident proident ipsum fugiat tempor eu cillum cupidatat. Aute voluptate do eu do laborum commodo est esse. Proident aute Lorem in fugiat nostrud cupidatat sit magna in occaecat. Aliqua ad id veniam deserunt consequat veniam fugiat eiusmod nisi excepteur eu exercitation qui ad.

Ea anim proident magna sunt in amet et tempor. Magna exercitation labore aliqua Lorem ipsum Lorem ut ipsum velit id labore eiusmod.

Incididunt ea aliquip nulla officia anim do et qui. Ullamco non labore minim tempor ad incididunt ut reprehenderit ea aute consectetur. Nulla quis cupidatat eiusmod irure anim quis qui. Sint ut aliqua et consequat officia duis aliquip dolore quis minim. Ipsum anim esse qui veniam ea aute ullamco incididunt ullamco duis. Eu sit dolor irure qui nostrud aute sit magna. Anim labore esse velit velit incididunt adipisicing ea fugiat cupidatat ex exercitation et. Est exercitation anim ea labore aliqua officia minim tempor excepteur irure id consectetur.

If it had real content, I would believe an article that looks like this.

Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
○	○	○	○	○	○	○

Review article and respond below.

Sit Cillum Labore Sunt Culpa Pariatur Fugiat

Cillum Eiusmod

Incididunt ut do velit ad id Lorem sint do et excepteur. Quis aute consequat officia veniam mollit ex sunt exercitation aute tempor dolor irure. Tempor sint consectetur non ut mollit dolore nostrud velit laborum.

Duis est dolore duis tempor sint sint aliquip esse culpa nostrud. Quis tempor sit cupidatat id dolor occaecat veniam eiusmod ipsum non aute. Officia et ex do officia ipsum amet commodo amet voluptate pariatur aliqua commodo sit aliquip. Excepteur dolore tempor laborum est eiusmod consectetur ipsum. In voluptate nisi nulla exercitation sint sunt velit commodo id quis.

Figure 2. Combinations of our factors' levels create different presentations. (*left*) High link density, a video positioned at the article's top, no images, a high word count, and small serif font. (*middle*) High link density, no video, no images, a low word count, and small sans serif font. (*right*) No links, no video, a high image count (not all shown), a medium word count, and large sans serif font. Images and videos were blurred with the CSS blur effect (20 px for images, 40 px for videos).

to "cat people" more than "dog people," let alone actual newsworthy topics like politics or economics. We recognize the tradeoff in making this choice, namely that we miss the possibility of detecting how presentational factors might *interact with* web page content. We leave that question for future work.

Pyrite displayed articles with specifically manipulated visual features. For text, it drew from Cicero's well-known 1st century B.C. "lorem ipsum" text [24], using either 348, 644, or 1070 words. Links were randomly applied according to a link density factor reflecting low, medium, and high density. Similarly, font sizes were chosen for title and body fonts that represented small, medium, and large sizes. Fonts were either serif or sans serif in style. (Again, for specific values, see Section 3.4, below.)

For images and videos, Pyrite used heavily blurred media that became indistinct so as not to distract participants with their content (Figure 2). Videos were positioned either at the top or in the middle of articles, but not at the bottom, as we did not observe that placement in our sample of 100 real news websites. Videos were under 60 seconds in duration, but all participants quickly realized that videos were meaningless, and ceased playing them.

3.3 Procedure

The experiment unfolded in three stages. First, we collected basic demographic data about our participants using an online questionnaire. (See Section 3.1, above.) Second, we used Pyrite to show participants a series of news-like articles, obtaining their perceived credibility ratings on a 7-point Likert scale, described in detail below. Third, we conducted post-session interviews in which we asked our participants questions about their experiences during the study.

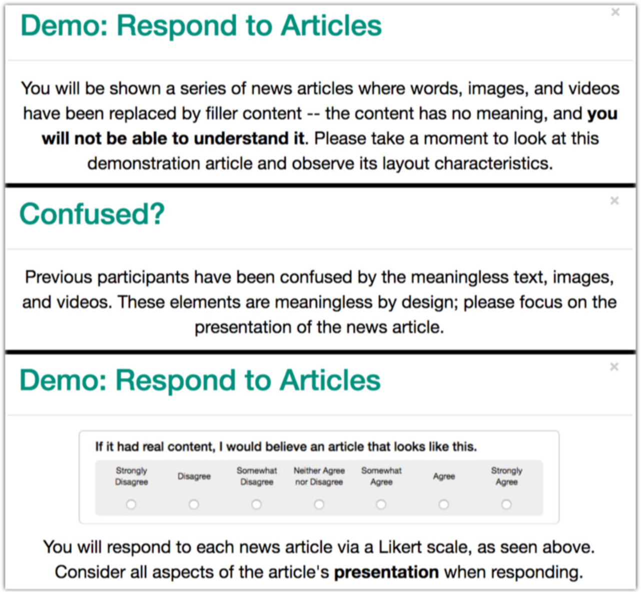
Participants were shown a randomized series of 24 web page "articles," some examples of which are shown in Figure 2. All participants provided responses for all 24 articles. For each article, participants responded on an agreement-based Likert scale ranging from 1 = "Strongly Disagree" to 7 = "Strongly Agree." The Likert scale appeared at the bottom of each Pyrite-generated web page. The prompt for the Likert scale was:

> *If it had real content, I would believe an article that looks like this.*

The wording of our prompt was influenced by prior work on web credibility [10], which discussed key terminology for investigating credibility, with the top three terms being "credible," "believable," and "reputable." We selected the word "believe" because of concerns that "credible" or "reputable" might lead participants more to consider the *authorship* or *provenance* of the article, rather than its mere *presentation*. Also, the word "believe" serves as an active verb for the participant.

In addition, our prompt was designed to avoid extracting mere professionalism judgments from our participants [39]. Our Pyrite-generated web pages, despite their controlled variations, all had a similar "look and feel." As a result, their level of professionalism was quite similar and unlikely to cause differences in perceived credibility. A more professionalism-oriented prompt might have focused on the niceness of the presentation, rather than on its believability. Our follow-up interviews gave no indication that professionalism was actually the underlying construct judged.

Prior to rating web page articles for their perceived credibility, participants did a training exercise that walked them through a practice article. Explanatory prompts were shown as participants scrolled through the practice article (Figure 3), and the experimenter verbally checked for participant understanding. Specifically, the first prompt (Figure 3, top) appeared when the page loaded; the second prompt (Figure 3, middle) appeared when the page had been scrolled down about one-third of the way; and the third prompt (Figure 3, bottom) appeared once the Likert scale at the bottom of the article became visible.



Demo: Respond to Articles

You will be shown a series of news articles where words, images, and videos have been replaced by filler content -- the content has no meaning, and **you will not be able to understand it**. Please take a moment to look at this demonstration article and observe its layout characteristics.

Confused?

Previous participants have been confused by the meaningless text, images, and videos. These elements are meaningless by design; please focus on the presentation of the news article.

Demo: Respond to Articles

If it had real content, I would believe an article that looks like this.

Strongly Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Strongly Agree
○	○	○	○	○	○	○

You will respond to each news article via a Likert scale, as seen above. Consider all aspects of the article's **presentation** when responding.

Figure 3. While scrolling through a practice article, participants were shown these three prompts. The prompts were shown one at a time; they are stacked here merely for considerations of space.

After rating all 24 articles presented to them, participants were asked to respond freely in a semi-structured interview to the following three questions:

1. “What are your first impressions of the pages you saw after completing this study?”
2. “How did you evaluate each page for its believability?”
3. “What elements or characteristics of each page did you find yourself looking at?”

Of course, we recognize that participants’ answers to these questions cannot be taken as “ground truth” for participants’ behavior. It is quite possible, and even likely, that participants did not know what they were looking at. Our intention, however, was to discover what participants *thought* mattered to them in making their credibility ratings, and what they *thought* they were looking at. Ultimately, ground truth for such questions lies beyond the scope of the current study; future work could pursue such answers with a different study design (*e.g.*, eye-tracking).

3.4 Design & Analysis

Recall that the levels of our factors were determined by our survey of 100 articles from 20 of the most popular U.S. news websites. (See Section 3.2, above.) Our study utilized a partial within-subjects factorial design with the following factors and levels administered to all participants in a fully-crossed design:

- *Video*: absent, present
- *Images*: 0, 3, 6
- *Link Density*: 0.000, 0.002, 0.007, 0.017 links per word
- *Trial*: 1-24

For each combination of the levels of the above primary factors, one level was selected randomly for each of the following secondary factors:

- *Words*: 348, 644, 1070
- *Font Size (body / title)*: 13/30, 16/38, 19/46 point
- *Font Face*: serif, sans serif
- *Video Placement*: top, middle of article

In addition to the above factors and levels, we examined participants’ *Age* and *Gender*. Values for *Age* ranged from 18-23 and four options were offered for *Gender*. (See footnote 1, above.)

We chose the above experiment design to allow us to investigate the presentational factors that we hypothesized might affect web page credibility based on prior work (*e.g.*, [7,9,13,20,41]), while also avoiding making our study impractical to run and analyze. (If *all* primary and secondary factors had been fully crossed, study sessions would have been impractically long.)

In all, 31 participants completed 24 trials for 744 total trials. A single “trial” was viewing a web page article generated by Pyrite and indicating a single credibility rating on the 1-7 Likert scale. The dependent variables for each trial were a *Credibility* ordinal response and *Page Time*, measured in milliseconds.

Our data analysis approach unfolded in stages. With many covariates, factors, and levels, we could not justifiably throw all potential effects into a single statistical model of high order. Instead, as is common practice, we followed a *factor screening* approach [29], whereby we first used exploratory data analysis, including descriptive statistics, graphical plots, and outlier analyses, to determine which factors seemed like they might exert a significant effect on either *Credibility* or *Page Time*. We also tested each factor in isolation and in all two-way interactions. Factors involved in statistically significant or trend-level results ($p < .10$) during this screening stage were preserved in the final statistical model. Factors that did not emerge as potentially significant were not explored further.

For the *Credibility* measure, the final statistical model had four fixed effects: *Video*, *Images*, *Words*, and *Font Size*. This meant that *Link Density*, *Font Face*, and *Video Placement* did not exert a detectable effect on *Credibility*, and were therefore dropped. *Video* was encoded as a dichotomous variable while *Images*, *Words*, and *Font Size* were encoded as ordinal variables corresponding to their respective low, medium, and high values. The statistical model also had *Trial* and *Subject* included as random effects [11,22].

For the *Page Time* measure, the final statistical model had three factors: *Video*, *Images*, and *Words*. This meant that *Link Density*, *Font Size*, *Font Face*, and *Video Placement* did not exert a detectable effect on *Page Time*, and were therefore dropped. Factor encodings were the same as for *Credibility*. Again, *Trial* and *Subject* were included as random effects.

We ran these statistical models according to established procedures. Specifically, we used the nonparametric Aligned Rank Transform [16,34,44] to analyze *Credibility* as an ordinal response. We utilized a parametric linear mixed model analysis of variance [11,22] for the log of *Page Time*, which was lognormally distributed [21]. Statistical tests were conducted in R using the `ARTool`, `lme4`, `car`, `phia`, and `emmeans` packages.

4 RESULTS

In this section, we present the results of our study of web page credibility perceptions based on web page presentational factors. We first discuss *Credibility* and then *Page Time*.

4.1 Perceived Credibility

Recall that participants rated 24 web page articles containing “lorem ipsum” text on 1-7 Likert scales ranging from 1 = “Strongly Disagree” to 7 = “Strongly Agree” in response to the prompt, “If it had real content, I would believe an article that looks like this.”

An omnibus test shows that there were significant main effects of *Video* ($F_{1,\,655.4} = 14.87$, $p < .001$), *Images* ($F_{2,\,654.8} = 13.33$, $p < .0001$), and *Font Size* ($F_{2,\,667.4} = 6.96$, $p < .01$) on *Credibility*. By contrast, *Words* did not exert a detectable main effect ($F_{2,\,667.3} = 2.06$, *n.s.*). However, there was a significant *Images* × *Words* interaction ($F_{4,\,665.7} = 2.43$, $p < .05$), and a significant *Images* × *Words* × *Font Size* interaction ($F_{8,\,664.9} = 2.19$, $p < .05$). In the following paragraphs, we discuss each of these effects in turn.

Figure 4 shows *Credibility* ratings for when video was absent or present. The mere presence of a video somewhat increased the perceived credibility of web page articles.

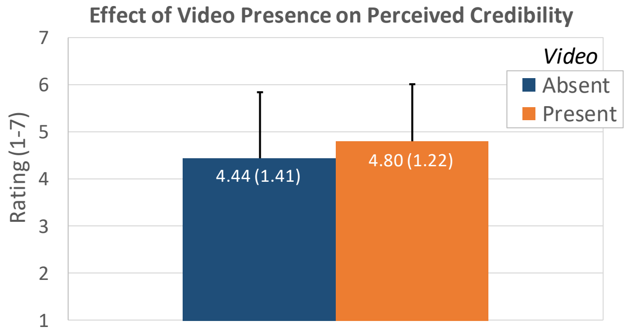


Figure 4. Average credibility ratings by video presence. Higher is “more credible.” Error bars are +1 *SD*.

Figure 5 shows *Credibility* ratings for each level of *Images*. Interestingly, it seems that three images might be more credible than either zero or six images. *Post hoc* pairwise comparisons using Tukey’s correction [42] indicate that zero images were significantly less credible than three or six images ($p < .01$), but that three and six images were not detectably different.

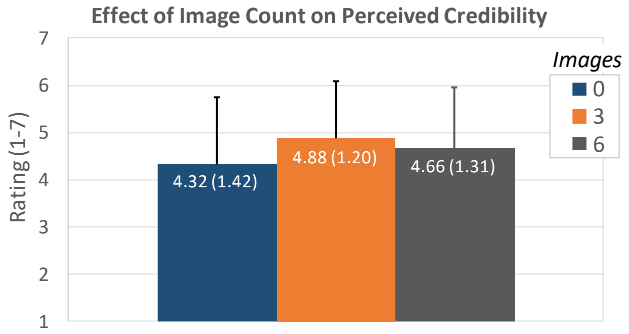


Figure 5. Average credibility ratings by image count. Higher is “more credible.” Error bars are +1 *SD*.

Recall, however, that *Credibility* was affected by a significant *Images* × *Words* interaction, as shown in Figure 6.

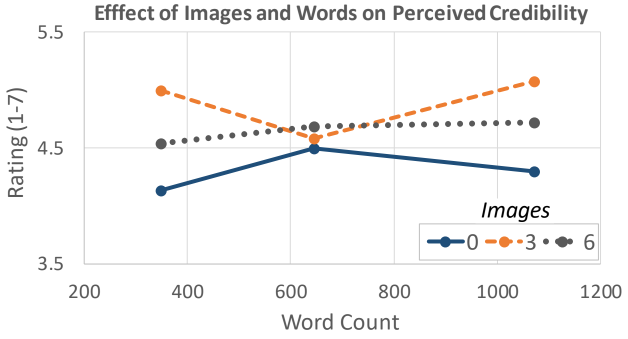


Figure 6. For 0, 3, and 6 images, credibility ratings changed over 348, 644, or 1070 words. Higher is “more credible.”

We can use interaction contrasts [25], corrected for multiple comparisons with Holm’s sequential Bonferroni procedure [17], to examine the credibility differences between image levels at 348, 644, and 1070 words. Results indicate that the significant difference in credibility between zero and three images at 348 words disappears at 644 words ($p < .05$). This difference does not quite re-emerge at 1070 words ($p = .15$). Therefore, the number of images used significantly affects perceived credibility when the word count is low, but no longer seems to matter as much when the word count increases. Articles with few words that have either no images or many images seem less credible.

Figure 7 shows *Credibility* ratings by level of *Font Size*. It seems that while small and medium font sizes had similar credibility ratings, large fonts reduced credibility. This result is confirmed by significant *post hoc* pairwise comparisons using Tukey’s correction [42], which show that large fonts were significantly less credible than both small or medium fonts ($p < .01$), but that small and medium fonts were not detectably different.

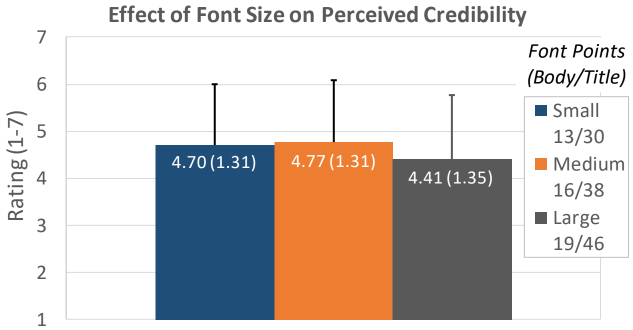


Figure 7. Average credibility ratings by font size. Higher is “more credible.” Error bars are +1 *SD*.

Recall, however, that *Font Size* was involved in a three-way interaction with *Images* and *Words*, adding nuance to the interaction present in Figure 6. Figure 8 shows the same plot as Figure 6, but now broken out by three levels of *Font Size*. It is clear that as *Font Size* changes, the *Images* × *Words* interaction also changes. Specifically, for small and large fonts, the medium word count brings credibility ratings for all three image levels together, but for medium fonts, credibility ratings at the medium word count diverge, converging instead more at the low word count. These observations are confirmed by significant interaction contrasts ($p < .05$).

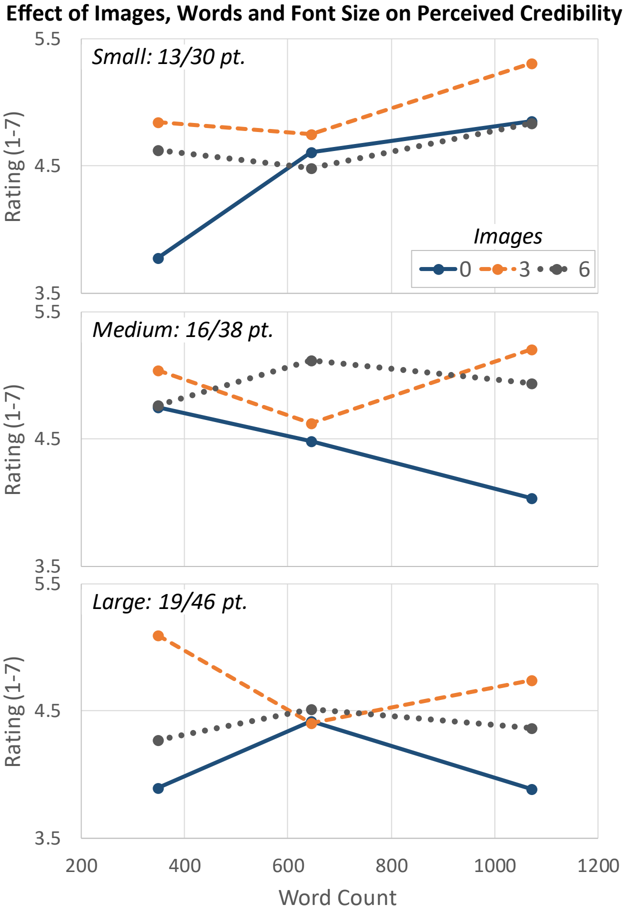


Figure 8. The *Images* × *Words* interaction over three levels of *Font Size*. Higher is "more credible."

4.2 Page Time

We also measured how long participants spent on each page while forming their credibility judgments, as page times might indicate the certainty or confidence of such judgments [40]. Faster times suggest that credibility judgments were easily formed, whereas slower times suggest more scrutiny was necessary. Also, basic sanity checking is available through an examination of page times.

An omnibus test showed that there were significant main effects of *Video* ($F_{1,\,676.5} = 9.16$, $p < .01$), *Images* ($F_{2,\,676.9} = 15.42$, $p < .0001$), and *Words* ($F_{2,\,683.9} = 10.02$, $p < .01$) on *Page Time*. There were no significant interactions among these factors. The following paragraphs discuss each effect in turn.

Figure 9 shows *Page Times* for when video was absent or present. The mere presence of an (unwatched) video increased time-on-page by just over one second on average.

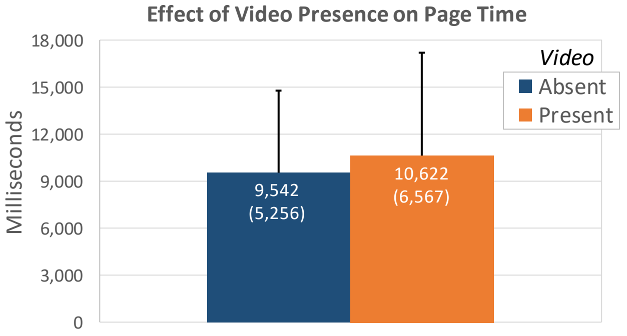


Figure 9. Average page times by video presence. Error bars are +1 *SD*.

Figure 10 shows *Page Times* for each level of *Images*. Expectedly, more images in an article resulted in more time spent, even when those images were content-free. *Post hoc* pairwise comparisons using Holm's sequential Bonferroni procedure [17] indicate that page times were significantly different among all three levels of *Images* ($p < .01$). It seems each image added about 350 ms on average.

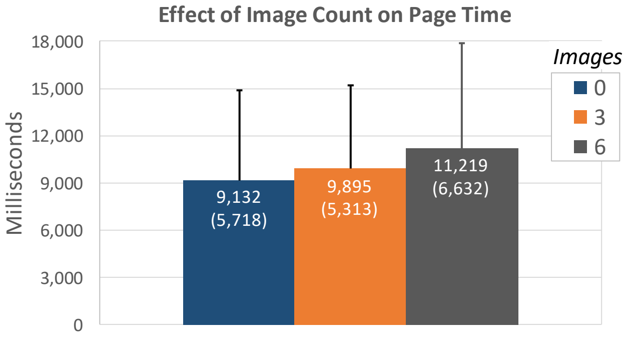


Figure 10. Average page times by number of images. Error bars are +1 *SD*.

Finally, Figure 11 shows how *Page Times* were affected by the number of "lorem ipsum" words. As with video and images, having more words resulted in longer page times. *Post hoc* pairwise comparisons using Holm's sequential Bonferroni procedure [17] indicate that many words produced significantly longer page times than either medium or few words ($p < .01$), but that medium and few word counts were not detectably different.

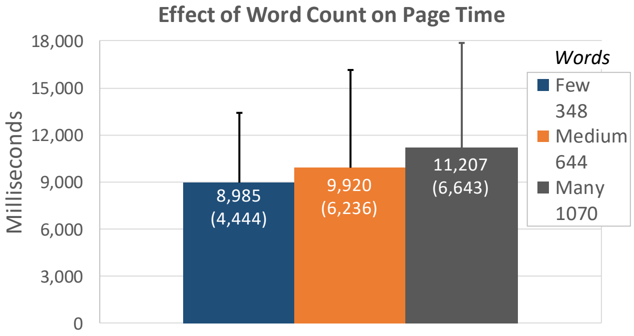


Figure 11. Average page times by word count. Error bars are +1 *SD*.

On the whole, then, it seems that times-on-page generally increased with increasing content (videos, images, and words). This result might seem unsurprising, but what is noteworthy here is that *the content was known by participants to be devoid of any meaning*. There was little, if anything, for participants to consume in the content itself. Nonetheless, participants were obligated to visually process web pages' stylistic features in order to decide upon its perceived credibility. And in that regard, page content—even meaningless content—seems to have played a role in forming credibility judgments.

The results in this section also allow us to assess the lighthearted claim that "a picture is worth a thousand words." In terms of page time, each image added about 350 ms (Figure 10). Similarly, every 100 words added about 310 ms (Figure 11). Thus, it seems that, at least in the case of heavily blurred images and "lorem ipsum" words, a picture is worth about 100 words.

4.3 Interview Results

Immediately after each participant completed the experiment, we conducted a semi-structured interview to discover qualitative insights about their experience. We asked participants about their first impressions, how they judged credibility, and which page elements they felt they noticed most. When appropriate, we asked follow-up questions to encourage participants to expound upon their responses. (See Section 3.3, above.)

Twelve of our 31 participants (39%) said that the *presence* of images or videos was the most impactful on their credibility ratings, and seven participants also commented on the *placement* of images and videos. Participant #11 said, "When there was a video in the middle of the article, I thought it looked less like a news article, since articles usually put the videos at the very top, since that's the main point."

Ten of our participants (32%) commented on font size being the factor they felt had the most impact on their credibility ratings, specifically that larger fonts lowered a page's perceived credibility. This sentiment is in agreement with our quantitative findings about large fonts and lower credibility (Figure 7), especially for short and long articles with no images (Figure 8).

Many statements participants made evidenced interactions between factors, such as when Participant #6 said, "If there was only a little text, and a lot of pictures, then I would find it less believable." This comment agrees with the quantitative finding that six images were rated as less credible than three images when both were in the presence of only a few words (Figure 6).

Overall, participants were clear that having some images and videos made articles seem more credible, but that having too many images reduced credibility. However, there was a lot of interaction between factors: most participants identified the impact of certain factors as being conditional on the levels of other factors, while some just said there needed to be a balance among levels of factors.

Other comments were quite specific about individual factors. For example, participant #14 said she was "looking at the how big the title was—you want to catch the person's attention, but at the same time, I feel like a lot of clickbait articles have huge titles because that's all they focus on, so I tried to see how big that was."

Thus, our interview results corroborated our experimental findings. Participants' evaluations of credibility indicated that certain factors had a bigger impact on credibility ratings than others—namely, video and image presence and placement, image count, and the interaction between image count, word count, and font size. We now discuss and reflect upon these results.

5 DISCUSSION

Our results indicate that the presence of videos and images affects credibility ratings, even apart from content. Simply having a video increased credibility from 4.44 to 4.80 on average (Figure 4). Having three images, as opposed to zero or six images, was also viewed as more credible (Figure 5), especially when word counts were low (Figure 6) and fonts were large (Figure 8). Overall, the smallest and medium font sizes were viewed as more credible than the largest font (Figure 7).

By comparison, link density, serif or sans serif fonts, and video placement did not affect perceived credibility. Perhaps these factors exerted no detectable influence on credibility because users were used to seeing such variations on credible websites. For example, *The New York Times* (nytimes.com) uses serif fonts, whereas *CNN* (cnn.com) uses sans serif fonts.

Unsurprisingly, times-on-page increased in the presence of videos, more images, and more words, even when those elements themselves were content-free. Also, our interview results agreed with our experimental findings, adding support to our discoveries.

Perhaps most interesting was the interaction between image count and word count (Figure 6). Specifically, at the lowest word count, having zero images was the least credible, six images was more credible, but three images was most credible. This pattern was exacerbated in the presence of large fonts (Figure 8). It seems as if participants were making a "Goldilocks judgment,"[3] where a moderate number of images suggested higher credibility.

Interestingly still, when word count increased from few (348) to medium (644) words, the "Goldilocks zone" of three images disappeared, with credibility mostly converging regardless of image count, except for medium font sizes. When word count increased further still to many (1070) words, judgments about image count began to resume their shape from 348 words, although to a slightly lesser degree. Thus, it seems when articles are of a medium length, the number of images is less crucial, but for short or long articles, the number of images matters more.

Our interviews also supported the idea of a "Goldilocks zone" with respect to words and images. Participants talked about credibility being a function of article length, neither too short nor too long—and length is a function not just of word count but also of other media like video and images. The old cliché that "the total

[3] Goldilocks is the protagonist in the famous 19th-century children's tale, *Goldilocks and the Three Bears*. Goldilocks looks for porridge that is neither too hot nor too cold, but "just right."

is more than the sum of its parts" seems at work here, where the overall gestalt of a page is responsible for participants' credibility judgments more than any single factor.

5.1 Design Implications

There are multiple implications for design from our study. For example, one could surmise that the *least* credible design would be one that has large fonts (19 pt. body font, 46 pt. title font), a short article length (348 words), no video, and no images. Similarly, the *most* credible design would have a small or medium font size (13 – 16 pt. body font, 30 – 38 pt. title font), a video, three images, and be of a longer length (1070 words). Of course, the exact values of these settings might vary somewhat from those we tested, but the general *direction* of these findings would presumably hold.

A concerning set of design implications pertains to people designing online news sites to deliberately propagate falsehoods. Those trying to increase the perceived credibility of their site's pages could conceivably make use of our findings to make those pages more believable. However, the same could be said for legitimate online news sources seeking to bolster their own credibility. Furthermore, by uncovering some of the purely presentational factors that affect credibility, citizens can become more aware of their perceptions and "look beyond the surface" to carefully scrutinize the credibility of their news sources.

5.2 Study Limitations

As with any study, ours had limitations. Numerous other presentational factors could have been included in our study but were not due to scope. Some of these other factors were whitespace amounts, color schemes, font families, image and video sizes, image placements, and additional web elements like charts, buttons, text boxes, checkboxes, and so on. The Web is a rich environment—isolating precisely which elements affect credibility is an ambitious undertaking.

Another limitation of our study was that only 31 college-aged students, or recent college graduates, were part of our participant pool. As stated in Section 3.1, above, this demographic ranks high among social media use and online news promulgation and consumption [3]. But our study findings might not generalize beyond this demographic. A fuller picture should be obtained by including other and more diverse participants, participants of different ages, different educational backgrounds, members of various political parties, and people from different geographies.

6 FUTURE WORK

Beyond addressing the study limitations raised above, future work on this topic is replete with interesting directions. With the knowledge gleaned from this study, we can begin to understand how aspects of visual appearance attract or repel the conveyance of credibility, independent of content.

The most obvious next step is further validation of our results, achievable through collection of data "in the wild," starting with articles that are known to be credible and examining their visual attributes to determine if they correlate with our findings. In a sense, this would be the reverse of our current study.

In addition, a promising further step could use our findings to apply distortions [38] to real news articles with actual content, validating that the distortions we make do indeed correlate with predicted credibility judgments.

Another study could, with participants' permission, examine articles linked in social media posts for their visual aspects, and the associated credibility attributed by the poster.

A limitation of our study was not knowing exactly where participants were looking and for how long. An eye-tracker would provide this information alongside credibility judgments and participants' subjective interview responses. Understanding how objective eye-tracking measurements might correspond with our findings would be valuable.

Lastly, our study was limited to the desktop or laptop computing environment, but many people today consume their news on smartphone or tablet devices. Replicating our study on such devices would yield an understanding of how device-independent these presentational factors are (or are not).

7 CONCLUSION

In this paper, we presented the results of a study investigating how purely presentational factors in content-free web page articles affect college students' credibility judgments. Unlike prior work on web credibility, which has focused on web content, or which has blended content with presentation, our work has *isolated* presentation by using "news articles" devoid of any meaningful content to assess the credibility of purely presentational factors. Our findings indicate that presentational factors, even isolated from content, *do* matter to perceived credibility. Specifically, video presence increased credibility, while large fonts and having no images reduced it. Having a few, but not too many, images increased credibility for short articles, especially in the presence of large fonts. Our subsequent interviews corroborated our quantitative experimental results, with participants saying they noticed font sizes and the presence of videos and images most when forming credibility judgments. This work has shed light on how online news credibility judgments are based not only on content but also on visual presentation. It is our hope that these findings can help inform both people and systems when making judgments about online news credibility.

ACKNOWLEDGMENTS

The authors thank the Informatics Capstone course advisors at the University of Washington. This work was supported in part by the University of Washington Information School and the Mani Charitable Foundation. Any opinions, findings, conclusions, or recommendations expressed in this work are those of the authors and do not necessarily reflect those of any supporter.

REFERENCES

[1] Hunt Allcott, Matthew Gentzkow. (2017). Social media and fake news in the 2016 Election. *Journal of Economic Perspectives 31* (2), pp. 211-236. https://doi.org/10.3386/w23089

[2] Jonathan Howard Amsbary, Larry Powell. (2003). Factors influencing evaluations of web site information. *Psychological Reports 93* (1), pp. 191-198. https://doi.org/10.2466/pr0.2003.93.1.191

[3] Anonymous. (2017). Social media fact sheet. *Pew Research Center*, January 12, 2017. Retrieved September 6, 2018 from https://pewrsr.ch/2jmwndT

[4] Ardion Beldad, Menno de Jong, Michaël Steehouder. (2010). How shall I trust the faceless and the intangible? A literature review on the antecedents of online trust. *Computers in Human Behavior 26* (5), pp. 857-869. https://doi.org/10.1016/j.chb.2010.03.013

[5] Nicholas C. Burbules. (1998). Rhetorics of the Web: Hyperreading and critical literacy. In *Page to Screen: Taking Literacy into the Electronic Era*. London, England: Routledge, pp. 102-122.

[6] Andrew J. Flanagin, Miriam J. Metzger. (2000). Perceptions of Internet information credibility. *Journalism & Mass Communication Quarterly 77* (3), pp. 515-540. https://doi.org/10.1177/107769900007700304

[7] Andrew J. Flanagin, Miriam J. Metzger. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society 9* (2), pp. 319-342. https://doi.org/10.1177/1461444807075015

[8] B. J. Fogg, Jonathan Marshall, Othman Laraki, Alex Osipovich, Chris Varma, Nicholas Fang, Jyoti Paul, Akshay Rangnekar, John Shon, Preeti Swani, Marissa Treinen. (2001). What makes web sites credible?: A report on a large quantitative study. *Proceedings of CHI 2001*. New York: ACM Press, pp. 61-68. https://doi.org/10.1145/365024.365037

[9] B. J. Fogg, Cathy Soohoo, David R. Danielson, Leslie Marable, Julianne Stanford, Ellen R. Tauber. (2003). How do users evaluate the credibility of web sites?: A study with over 2,500 participants. *Proceedings of DUX 2003*. New York: ACM Press, pp 1-15. https://doi.org/10.1145/997078.997097

[10] B. J. Fogg, Hsiang Tseng. (1999). The elements of computer credibility. *Proceedings of CHI 1999*. New York: ACM Press, pp. 80-87. https://doi.org/10.1145/302979.303001

[11] Brigitte N. Frederick. (1999). Fixed-, random-, and mixed-effects ANOVA models: A user-friendly guide for increasing the generalizability of ANOVA results. In *Advances in Social Science Methodology*, Bruce Thompson (ed.). Stamford, Connecticut: JAI Press, pp. 111-122. http://eric.ed.gov/?id=ED426098

[12] Krisandra S. Freeman, Jan H. Spyridakis. (2004). An examination of factors that affect the credibility of online health information. *Technical Communication 51* (2), pp. 239-263. https://bit.ly/2VozRQi

[13] Susanne Furman. (2009). Credibility. *Usability.gov*, October 1, 2009. Retrieved September 6, 2018 from https://bit.ly/2L9sszC

[14] Jeffrey Gottfried, Elisa Shearer. (2017). American's online news use is closing in on TV news use. *Pew Research Center*, September 7, 2017. Retrieved September 12, 2017 from https://pewrsr.ch/2wKitZ0

[15] R. G. Hass (1981). Effects of source characteristics on cognitive response and persuasion. In *Cognitive Responses in Persuasion*, R. E. Petty, T. M. Ostrom, and T. C. Brock (eds.). Hillsdale, New Jersey: Lawrence Erlbaum, pp. 141-172.

[16] James J. Higgins, Suleiman Tashtoush. (1994). An aligned rank transform test for interaction. *Nonlinear World 1* (2), pp. 201-211.

[17] Sture Holm. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics 6* (2), pp. 65-70. http://www.jstor.org/stable/4615733

[18] Elle Hunt. (2016). What is fake news? How to spot it and what you can do to stop it. *The Guardian*, December 17, 2016. Retrieved September 16, 2017 from https://www.theguardian.com/media/2016/dec/18/what-is-fake-news-pizzagate

[19] Mineabere Ibelema, Larry Powell. (2001). Cable television news viewed as most credible. *Newspaper Research Journal 22* (1), pp. 41-51. https://doi.org/10.1177/073953290102200104

[20] Jinwoo Kim, Jae Yun Moon. (1998). Designing towards emotional usability in customer interfaces—trustworthiness of cyber-banking system interfaces. *Interacting with Computers 10* (1), pp. 1-29. https://bit.ly/2XAmn5X

[21] R. J. Lawrence. (1988). The log-normal as event-time distribution. In *Log-normal Distributions: Theory and Application*, E. L. Crow and K. Shimizu (eds.). New York: Dekker, pp. 211-228.

[22] Ramon C. Littell, P. R. Henry, C. B. Ammerman. (1998). Statistical analysis of repeated measures data using SAS procedures. *Journal of Animal Science 76* (4), pp. 1216-1231. https://doi.org/10.2527/1998.7641216x

[23] Ziming Liu. (2004). Perceptions of credibility of scholarly information on the Web. *Information Processing & Management 40* (6), pp. 1027-1038. https://doi.org/10.1016/S0306-4573(03)00064-5

[24] Lorem Ipsum. (2017). *Wikipedia*. Accessed on April 23, 2019. https://en.wikipedia.org/wiki/Lorem_ipsum

[25] Leonard A. Marascuilo, Joel R. Levin. (1970). Appropriate post hoc comparisons for interaction and nested hypotheses in analysis of variance designs: The elimination of Type IV errors. *American Educational Research Journal 7* (3), pp. 397-421. https://www.jstor.org/stable/1161635

[26] Marc Meola. (2004). Chucking the checklist: A contextual approach to teaching undergraduates web-site evaluation. *Libraries and the Academy 4* (3), pp. 331-344. https://doi.org/10.1353/pla.2004.0055

[27] Miriam J. Metzger, Andrew J. Flanagin. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics 59* (B), pp. 210-220. https://doi.org/10.1016/j.pragma.2013.07.012

[28] Miriam J. Metzger, Andrew J. Flanagin, Keren Eyal, Daisy R. Lemus, Robert M. Mccann. (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* (1), pp. 293-335. https://doi.org/10.1080/23808985.2003.11679029

[29] Max D. Morris. (2006). An overview of group factor screening. In *Screening*, Angela Dean and Susan Lewis (eds.). New York: Springer, pp. 191-206. https://doi.org/10.1007/0-387-28014-6_9

[30] Johan Leif Olaisen. (1990). Information quality factors and the cognitive authority of electronic information. In *Information Quality: Definitions and Dimensions*, Irene Wormwell (ed.). London, England: Taylor Graham, pp. 91-121.

[31] Thorsten Quandt. (2008). (No) news on the World Wide Web? *Journalism Studies 9* (5), pp. 717-738. https://doi.org/10.1080/14616700802207664

[32] David Robins, Jason Holmes. (2008). Aesthetics and credibility in web site design. *Information Processing & Management 44* (1), pp. 386-399. https://doi.org/10.1016/j.ipm.2007.02.003

[33] Paul I. Rosenthal. (1971). Specificity, verifiability, and message credibility. *Quarterly Journal of Speech 57* (4), pp. 393-401. http://www.tandfonline.com/doi/abs/10.1080/00335637109383084

[34] K. C. Salter, R. F. Fawcett. (1985). A robust and powerful rank test of treatment effects in balanced incomplete block designs. *Communications in Statistics: Simulation and Computation 14* (4), pp. 807-828. https://doi.org/10.1080/03610918508812475

[35] Craig Silverman. (2016). This analysis shows how viral fake election news stories outperformed real news on Facebook. *Buzzfeed*, November 16, 2016. Retrieved September 12, 2017 from https://bzfd.it/2f5IngQ

[36] Craig Silverman, Jeremy Singer-Vine. (2016). Most Americans who see fake news believe it, new survey says. *Buzzfeed*, December 6, 2016. Retrieved September 12, 2017 from https://bzfd.it/2gazWOR

[37] Michael D. Slater, Donna Rouner. (1996). How message evaluation and source attributes may influence credibility assessment and belief change. *Journalism & Mass Communication Quarterly 73* (4), pp. 974-991. https://doi.org/10.1177/107769909607300415

[38] Brendan Spillane, Séamus Lawless, Vincent Wade. (2017). Perception of bias: The impact of user characteristics, website design and technical features. *Proceedings of WI 2017*. New York: ACM Press, pp. 227-236. https://doi.org/10.1145/3106426.3106474

[39] Jacklin Stonewall, Michael C. Dorneich. (2016). A process for evaluating the gender and professionalism of web design elements. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting 60* (1), pp. 750-754. https://doi.org/10.1177/1541931213601172

[40] Noam Tractinsky, Avivit Cokhavi, Moti Kirschenbaum, Tal Sharfi. (2006). Evaluating the consistency of immediate aesthetic perceptions of web pages. *International Journal of Human-Computer Studies 64* (11), pp. 1071-1083. https://doi.org/10.1016/j.ijhcs.2006.06.009

[41] Alexandre N. Tuch, Eva E. Presslaber, Markus Stöcklin, Klaus Opwis, Javier A. Bargas-Avila. (2012). The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments. *International Journal of Human-Computer Studies 70* (11), pp. 794-811. https://doi.org/10.1016/j.ijhcs.2012.06.003

[42] John W. Tukey. (1949). Comparing individual means in the analysis of variance. *Biometrics 5* (2), pp. 99-114. http://www.jstor.org/stable/3001913

[43] C. Nadine Wathen, Jacquelyn Burkell. (2001). Believe it or not: Factors influencing credibility on the Web. *Journal of the Association for Information Science and Technology 53* (2), pp. 134-144. https://doi.org/10.1002/asi.10016

[44] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, James J. Higgins. (2011). The Aligned Rank Transform for nonparametric factorial analyses using only ANOVA procedures. *Proceedings of CHI 2011*. New York: ACM Press, pp. 143-146. https://doi.org/10.1145/1978942.1978963