The impact of readability on trust in information

Permalink
https://escholarship.org/uc/item/24c472qm

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN
1069-7977

Authors
Withall, Amanda
Sagi, Eyal

Publication Date
2021

Peer reviewed
The impact of readability on trust in information

Amanda Withall (amandawithall@stfrancis.edu)
Eyal Sagi (esagi@stfrancis.edu)
Department of Psychology, University of St. Francis
500 Wilcox St., Joliet, IL 60435, USA

Abstract

The increased prevalence of “fake news” in recent years makes it important for us to better understand how we decide which information to trust and accept as true. Existing research has identified the important roles of factors such as trust in the source of the information and reading ability. We report on the results of a study on how the style of writing, and specifically the difficulty level of the text providing the information, affects our trust in the information provided. Participants trusted the information more in texts that were more difficult to read, regardless of whether the information presented was correct or incorrect. Regardless of readability, participants trusted paragraphs providing truthful information more than those providing false information. This result has important implications for information on social media platforms, where the source may or may not be readily available and is often not an expert on the topic.

Keywords: fake news; false information; deception detection; readability

Introduction

The rapid development of blogs, social media, and news websites is bringing a new, and mostly uncharted, era of information sharing. Lightning-fast streams, available at our fingertips, have fundamentally changed how we absorb information. Nevertheless, as the speed at which information is disseminated has dramatically increased, the overall quality of it has apparently declined. False news has been shown to spread "farther, faster, deeper, and more broadly" than verified news stories online (Vosoughi, Roy, and Aral, 2018).

Furthermore, it seems that readers generally treat false information differently than true information. Mendoza, Poblete, and Castillo (2010) conducted a case study of information spread over Twitter during the 2010 Chilean earthquake. In this study, rumors and misinformation were interacted with – and questioned – more than verified news articles. Reassuringly, this seems to indicate more skepticism towards false information.

Despite this, people accept "fake news" as true fairly often. The concept of truth bias offers one explanation for this. We are predisposed to believe that others are telling the truth until otherwise perceived (see Street and Masip, 2015 and Levine, 2014). As consumers of false information that is presented factually, this means that we are often not skeptical enough of it.

Other lines of research have demonstrated that the belief system of an individual acts as a mediator for accepting false information. Druckman, Peterson, and Slothuss (2013) were able to show that participants had the ability to discriminate between strong arguments (containing fact and objective reasoning) and weak arguments (substantially lacking in fact) because they generally adjusted their opinions in favor of the stronger argument. That is, until the arguments are given polarizing party endorsements. When polarized party endorsements where presented, the political party their participant identified with became the determining factor of persuasion for the given information, even when the argument for the participants' side was weak.

Most factors that have been shown to affect the acceptance of false information as true can be considered external factors as they are not directly related to the information provided. One of the most recognizable of these factors is source credibility. The effect that the perceived credibility of a source has on the acceptance of information provided by it has been extensively studied (e.g., Sternthal et al., 1978; Pornpitakpan, 2004). The effect of perceived expertise, a more specific type of source credibility, has also been examined in numerous studies. This includes studies on the persuasiveness of a text (Tormala et al., 2006) and on how the expert's agreement with the reader's beliefs impacts the reader's perception of accuracy (Clark et al., 2012).

Other external factors that have been identified by research can be connected to processing fluency effects. At the core of these effects is the observation that readers judge a given piece of information not only by its content, but by how easily they can understand it. Higher levels of fluency (ease of understanding) positively correlate with trust in the information, its likeability (see Wänke and Hansen, 2015), as well as the formation of beliefs based on it (see Reber and Greifender, 2017).

Fluency effects can come about in a variety of ways. For example, the illusory truth effect – where repeated statements cause us to perceive information as more truthful even when it is not – can be attributed to fluency through familiarity. Boehm (1994) demonstrates this by showing that the illusory truth effect is mediated by familiarity. Pennycook et al. (2018) study this effect in online settings and show that even a handful of exposures to fake news titles are enough to raise their perceived correctness. More importantly, Fazio et al. (2015) provide evidence that these effects persist even if...
participants have prior knowledge that falsifies the given information. Fluency effects therefore appear to exert a substantial influence over judgments of trust, even for false information.

Fictional narratives provide another case in which fluency effects may elicit undue trust. Narratives providing context to misinformation likely improve the reader’s understanding and consequently their trust in the content, even when they have prior knowledge that falsifies the claims (Hamby et al., 2020). In fact, narratives that contain false information that is presented as true frequently lead readers to accept the truthfulness of the information (Appel and Richter, 2007), most likely because the context adds a new level of understanding.

While there is a substantial body of research on the effect of factors that are external to the text on trust in its content, the influence of factors that are inherent to the text is not as clear. Such factors may include features of the text or information itself, such as word choice or sentence length as well as syntactic complexity and the degree to which the text relies on pragmatic inference. Gilbert, Krull, and Malone (1990) report that simply being able to comprehend false information increases the likelihood that it will be viewed as true later, as you might expect from biases rooted in processing fluency. They also go on to suggest that it seems likely that readability and grammatical features, such as the tenses used, could similarly affect readers’ susceptibility to false information. Some twenty-five years later, Martin, Summerville, and Wickline (2017) explored the idea that text difficulty interacts with source credibility to influence attitudes towards a text but were unable to substantiate it. This may have been due to the differences in reading levels used – their manipulation relied on texts that were one grade level apart on the Flesch-Kincaid scale, and both levels would be considered difficult for the average American reader. More promising results for the possible effect of textual features on trust come from a study on privacy policies. Ermakova, Baumann, Fabian, and Krasnova (2014) found that higher subjective readability ratings correlated with higher trust ratings.

Aside from these studies, the topic has not been directly addressed in psychological literature. However, recent research in information technology and computer science provides some interesting insight to the topic. Santos et al. (2020) found that readability features have a significant impact on automated fake news detection. Likewise, Rubin et al. (2015) used rhetorical structure theory to examine differences between true and false news stories to see if it was possible to detect either one using it, and found varying degrees of success, depending on the individual feature. Chen et al. (2015) proposes that textual features such as vocabulary level and pronoun utilization may be correlated with identifying ‘clickbait’ articles and used for their detection.

The fact that there are definable differences in the textual features of true and false information, mixed with the idea that humans are inclined to trust information they understand or like, makes studying factors inherent to the text important. To better understand the effects that false information has on the general population, we need to examine how the features that make it different impact our perceptions of trust. That is the purpose of the present paper.

We examine how the reading level of a text affects perceptions of its truthfulness, even when the information presented is false. We believe that the readability of a text will influence the perceived truthfulness, similar to how it impacts a machine’s ability to do the same task. We believe that increased difficulty will affect participants’ ability to detect false information due to the increased demands on cognitive resources used to comprehend the information. We hypothesize that this will lead to similar levels of trust in difficult texts regardless of whether they contain correct or incorrect information. In contrast, we predict that easier texts will show a discrepancy in trust between correct and incorrect texts, where incorrect texts are trusted less than correct ones, because the increased readability will allow participants to discern between correct and incorrect information.

In our experiment, participants read passages that present correct or incorrect information. These passages are written at either a high or low reading level, as measured by the Flesch-Kincaid scale. We hypothesize that incorrect information will be easier to detect as part of a low reading level passage. As a result, we predict a larger effect of correctness for lower reading level passages than high reading level ones. We further hypothesized that participants will have an easier time reading the more readable passages which will enable them to further scrutinize them. Because of this added scrutiny, we predict that participants will trust the low reading level passages less. To assist with interpreting the results, two additional measures (likeability and perceived difficulty) were collected about passages.

**Experiment 1**

**Participants** 55 participants were recruited through Amazon’s Mechanical Turk (see Casler, Bickel, and Hackett, 2013). Twelve additional participants were not counted due to voluntarily exiting the study before submission or timing out of the study. All participants were over the age of eighteen.

**Materials** We selected four major categories – politics, economics, technology, and medicine – that news media sites and profiles frequently discuss, and then selected eight subtopics from each major category for a total of thirty-two writing topics. A short informational passage (word count $M = 98$) was written for each topic and then adapted into each variable instance, resulting in 128 passages total. We determined reading level using Flesch-Kincaid, with high level paragraphs falling between grade levels eleven and thirteen ($M = 12.566$) and low-level paragraphs falling between grade levels five and seven ($M = 7.214$). In each correct passage, 3 phrases were changed to create the incorrect version of the passage. These differences are similar in both reading level versions of each passage. Figure one displays a sample of the materials used.
High reading level passage (correct/incorrect)

Cancer is when abnormal/normal cells begin to grow and develop in a specific region of the body and can display in a wide variety of symptoms depending on where it is located. The most common causes of cancer are tobacco use, obesity, and exposure to disease. Cancers can begin to migrate to other parts of the body if not treated with options such as surgery, radiation therapy, or chemotherapy, where rounds of medications are administered to halt/encourage disease development. The death rate for cancer across the world is fairly high/low.

Low reading level passage (correct/incorrect)

Cancer happens when cells that did not grow/grow normally start to take over a part of someone's body. There are many kinds and signs of cancer because it can affect many different places in our body. The things that cause cancer the most are smoking, body weight, and diseases. Cancer in one part of the body often spreads to other parts if it is not treated. Cancer can be treated with chemotherapy, which is when someone takes a few different drugs to try to stop the disease. Other treatments, such as surgery or radiation, are also/rarely useful. Many/Only a few people die from cancer all over the world.

Procedure The study was conducted using the FindingFive online platform (FindingFive Team, 2019). Participants were initially told that the study was to determine how reading level impacts their opinions about a text. Each participant completed two practice paragraphs to teach them how to use the interface. Then, every participant was exposed to all thirty-two topics, with conditions (2 (high reading level or low reading level) x 2 (containing correct information or incorrect information) for each topic randomized and counterbalanced across participants. After each paragraph, participants were asked three five-point Likert scale questions to determine their level of trust in the given information, how much they liked or disliked the passage, and how difficult they felt the topic was to understand. Attention checks were given every eight passages.

Results and Discussion

We averaged the ratings of each of the three questions on a per-participant basis. Table 1 presents the overall means for each question by condition.

A generalized linear model with participant and passage as random effects and readability as a dependent variable revealed that participants found the correct versions as more readable than the incorrect versions ($F(1, 1671) = 39.37, p < .0001$). As expected, more difficult texts were also judged as less readable than easier texts ($F(1, 1671) = 32.46, p < .0001$). There was no interaction observed between the two variables ($F(1, 1671) < 1, n.s.$). A similar analysis with likeability as the dependent variable showed similar results. Participants liked the correct texts more than incorrect texts ($F(1, 1671) = 268.47, p < .0001$). Perhaps surprisingly, participants liked the more difficult texts better than the lower difficulty texts ($F(1, 1671) = 9.22, p < .0001$). As with the analysis of readability ratings, the interaction was not statistically significant ($F(1, 1671) = 1.86, p = .17$).

The key analysis in this study involved the trust participants report in the texts after reading them. As before, we conducted a similar generalized linear model for this purpose. Unsurprisingly, participants overwhelmingly preferred the correct texts over the incorrect ones ($F(1, 1671) = 300.94, p < .0001$). More importantly, the also trusted the more difficult texts more than the less difficult ones ($F(1, 1671) = 23.87, p < .0001$). In contrast with our prediction, the interaction between correctness and difficulty was not significant ($F(1, 1671) = 3.41, p = .065$).

Because participant’s ratings of subjective readability are more indicative of their actual experience of reading the passage, we also conducted an additional post-hoc analysis using that variable instead of the object text difficulty variable. This post-hoc analysis found a similar pattern of results, with one key difference – the interaction between readability and correctness was statistically significant ($F(1, 1671) = 5.16, p = .023$). Finally, the ratings of trust and readability for each passage were significantly correlated ($r(126) = .441, p < .0001$). Figure 2 depicts this relationship between ratings of trust and readability.

It is also possible that a participants’ reading level affects how they perform in this task. While we do not have a direct

Table 1: Mean ratings by participants in each condition.

<table>
<thead>
<tr>
<th>Text Difficulty</th>
<th>Correctness</th>
<th>Readability (SD)</th>
<th>Likeability (SD)</th>
<th>Trust (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Correct</td>
<td>3.48 (0.74)</td>
<td>3.30 (0.53)</td>
<td>3.72 (0.66)</td>
</tr>
<tr>
<td></td>
<td>Incorrect</td>
<td>3.20 (0.77)</td>
<td>2.79 (0.64)</td>
<td>2.98 (0.83)</td>
</tr>
<tr>
<td>Low</td>
<td>Correct</td>
<td>3.68 (0.67)</td>
<td>3.23 (0.50)</td>
<td>3.57 (0.66)</td>
</tr>
<tr>
<td></td>
<td>Incorrect</td>
<td>3.45 (0.70)</td>
<td>2.60 (0.73)</td>
<td>2.67 (0.85)</td>
</tr>
</tbody>
</table>
measure of reading ability, we can use the difficulty ratings as a proxy for this measure. Participants who find the texts easy are likely to be better readers than those who find them more difficult. We conducted a median split on participants based on their mean readability score and conducted the same general linear model above on each group. While the main effects did not change, the interaction between text correctness and text difficulty was significant for the group of participants who reported the texts as more easy to read ($F(1, 803) = 4.43, p = .036$) but not for the group who reported more difficulty with the texts ($F(1, 834) < 1, n.s.$).

Taken as a whole, our results provide a more complex answer to the hypothesis than initially expected. From our primary analysis, we did not find support for the first part of the hypothesis – participants reliably trusted correct information more than incorrect information across both correctness levels. There was, however, support for the second part of the hypothesis: paragraphs with a low objective readability show higher trust ratings for correct passages versus incorrect ones.

This discrepancy prompted us to take a deeper look at the data. In accordance with our results, prior research has indeed supported the idea that subjective readability ratings correlated to trust in texts more so than objective readability ratings (e.g., Ermakova et al., 2014). This may indicate that the difficulty of texts is judged not only on features such as word length, but also on how well it integrates with the current knowledge base of the reader. This interpretation of the results essentially suggests that the effect we observed represents another way in which processing fluency affects people’s judgment of trust.

Surprisingly, and in contrast to Ermakova et al. (2014), our participants rated the objectively more difficult texts as more trustworthy overall. This effect is explained by the fact that our participants rated correct texts as more readable than incorrect ones, also in line with processing fluency effects. However, it also led us to believe that the perceived readability score is not independent of the trust rating provided by the participants. This is supported by a modest, but statistically significant positive correlation between them ($r(2023) = 0.318, p < .001$). It is possible that participants noted each inconsistency between the text and their prior knowledge and worked to reconcile them. This deliberative process impacted their trust ratings accordingly (see Bago, et al., 2020) but also lowered their perception of how easily they could understand the text, leading to a higher difficulty rating.

Another possible explanation is that features of the text itself are used to make assumptions about external factors like the trustworthiness of the source, and then veracity judgements are made based on these assumptions. In the absence of any other indicator, participants relied on overall reading difficulty as proof that it came from a more credible source. Therefore, they trusted difficult texts more than easier ones in each correctness condition, even when both texts

Figure 2: Scatterplot of trust and difficulty ratings for each passage. Solid lines represent regression lines for correct passages. Dashed lines represent regression lines for incorrect passages.
Table 2: Mean ratings by participants in each condition.

<table>
<thead>
<tr>
<th>Text Difficulty</th>
<th>Correctness</th>
<th>Expertise</th>
<th>Readability (SD)</th>
<th>Likeability (SD)</th>
<th>Trust (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Correct</td>
<td>Expert</td>
<td>3.49(0.96)</td>
<td>3.33(0.87)</td>
<td>4.04(0.53)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-expert</td>
<td>3.38(0.94)</td>
<td>2.98(0.94)</td>
<td>2.97(0.85)</td>
</tr>
<tr>
<td>High</td>
<td>Incorrect</td>
<td>Expert</td>
<td>3.53(0.99)</td>
<td>3.24(0.87)</td>
<td>3.93(0.53)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-expert</td>
<td>3.25(1.04)</td>
<td>2.86(0.83)</td>
<td>2.67(0.89)</td>
</tr>
<tr>
<td>Low</td>
<td>Correct</td>
<td>Expert</td>
<td>3.79(0.90)</td>
<td>3.35(0.93)</td>
<td>3.95(0.58)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-expert</td>
<td>3.62(0.94)</td>
<td>3.10(0.93)</td>
<td>3.02(0.89)</td>
</tr>
<tr>
<td>Low</td>
<td>Incorrect</td>
<td>Expert</td>
<td>3.51(0.91)</td>
<td>3.07(1.01)</td>
<td>3.59(0.69)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-expert</td>
<td>3.33(1.02)</td>
<td>2.88(0.96)</td>
<td>2.69(0.86)</td>
</tr>
</tbody>
</table>

contained the same incorrect information. Experiment 2 tests this hypothesis – that participants use reading level as a surrogate for source expertise.

**Experiment 2**

Experiment 2 used the same materials and overall procedure as experiment one; however, the names of “authors” and their perceived area of expertise were attached to passages. The perceived area of expertise was manipulated by indicating the profession of the author (e.g., ‘family physician, ‘minister’, etc.).

**Participants** 52 participants were recruited from an introductory psychology course at a small Midwestern university. Four participants were eliminated from the participant pool due to failing attention checks or abandoning the study before completion. All participants were over the age of eighteen.

**Procedure** Experiment 2 followed the same procedure as Experiment 1. The text presented to participants included the author’s name an profession at the bottom right of the screen, following the text. As a result, the expertise of the source was an additional independent variable. The study therefore had a 2x2x2 design with 8 total conditions. Each participant saw 4 passages from each condition and the assignment of passages to conditions was randomized across participants.

**Results**

We again averaged the ratings of each of the three questions on a per-participant basis. Table 2 presents the overall means for each question by condition. As in Study 1, there were significant positive correlations between the ratings provided by participants: difficulty and likeability ($r(1568) = .440, p < .0001$), difficulty and trust ($r(1568) = .352, p < .0001$), and trust and likeability ($r(1568) = .493, p < .0001$).

Our main analysis concerns the ratings of trust in each condition and used a generalized linear model with participants and passages as random effects. Again, we found that participants trusted correct texts over incorrect ones ($F(1, 1478) = 16.677, p < .001$). However, we did not find a significant effect of text difficulty on reported trust ($F(1, 1478) = 0.051, p=0.821$). This was instead replaced by a higher level of trust in experts versus non-experts ($F(1, 1478) = 48.947, p < .001$). There was again no interaction between correctness and reading level ($F(1, 1478) = 2.959, p =0.086$) nor was there an interaction between correctness and expertise ($F(1, 1478) = 0.028, p =0.868$). A small but significant interaction was found between reading level and expertise ($F(1, 1478) = 4.053, p =0.044$). No three-way interaction was found between the variables ($F(1, 1478) = 0.276, p =0.600$).

**Discussion**

In experiment 1 we observed a significant effect of text difficulty on trust. When we included information about the author in study 2, this effect disappeared and was replaced by an effect of subject matter expertise. This supports our hypothesis that participants were using text difficulty as a cue for the subject matter expertise of the source. When the source of a text is not provided, we believe readers use other contextual information as cues to the trustworthiness of the source (E1). When the source is known, readers no longer need to rely on such contextual cues (E2).

However, this does not mean that readers completely ignored these cues, such as the difficulty of a text. Our results show an interaction between reading level and expertise, where participants trusted a high difficulty text from an expert author more than a low difficulty text from the same author but text difficulty did not influence participants’ judgments of trust when the author was not a subject-matter expert.

This suggests that participants expect experts to write at a higher level and are even more willing to trust the given information when those expectations are fulfilled. It also provides some evidence for a hypothesis by Martin, Summerville, and Wickline (2017) which they were unable to substantiate in their own studies: that text difficulty
interacts with source expertise to influence how people feel about a text.

Importantly, in this paper we present evidence that features of a text – here, text difficulty as measured by word and sentence length – impact the trust readers place in information they read. However, these effects are likely mediated by other cognitive factors.

In our studies, text difficulty appears to act as a cue for source expertise. It led to higher levels of trust when a source is not known (E1). However, when the source of the text was stated to be a subject-matter expert, a higher-than-average text difficulty reinforced the idea that the source is credible and increases trust further (E2). The accuracy of texts remained an overwhelmingly significant main effect regardless of other factors, indicating that participants in our studies were generally able to tell the difference between true and false information. The results draw a distinct connection between text features, external factors, and information veracity that goes beyond what we might expect from processing fluency effects alone.

It is important to point out that our materials were both diverse in topic and exceptionally wrong in incorrect instances. Some topics, such as cancer, were probably familiar to participants and appealed to prior knowledge more than topics like bionics. Importantly, although we did not control for prior knowledge in these studies, we believe that it is an additional important factor that is likely to interact with the effects we report here. Future studies could help assess this possibility.

References


