The Effects of User Comments on Science News Engagement

SPENCER WILLIAMS and GARY HSIEH, University of Washington, USA

Online sources such as social media have become increasingly important for the proliferation of science news, and past research has shown that reading user-generated comments after an article (i.e. in typical online news and blog formats) can impact the way people perceive the article. However, recent studies have shown that people are likely to read the comments before the article itself, due to the affordances of platforms like Reddit which display them up-front. This leads to questions about how comments can affect people's expectations about an article, and how those expectations impact their interest in reading it at all. This paper presents two experimental studies to better understand how the presentation of comments on Reddit affects people's engagement with science news, testing potential mediators such as the expected difficulty and quality of the article. Study 1 provides experimental evidence that difficult comments can reduce people's interest in reading an associated article. Study 2 is a pre-registered follow-up that uncovers a similarity heuristic; the various qualities of a comment (difficulty, information quality, entertainment value) signal that the article will be of similar quality, ultimately affecting participants' interest in reading it. We conclude by discussing design implications for online science news communities.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Additional Key Words and Phrases: comments, reddit, r/science, science news, science communication

ACM Reference Format:
Spencer Williams and Gary Hsieh. 2021. The Effects of User Comments on Science News Engagement. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 32 (April 2021), 29 pages. https://doi.org/10.1145/3449106

1 INTRODUCTION

The landscape of science journalism has changed dramatically over the past decade [48]. With the decline of traditional newspapers in the US and beyond, the already-fringe role of science journalists in traditional news outlets has decreased further, pushing them into social media and other online channels in order to remain active and visible to the public [11]. This emphasis on social media means that science news now frequently appears in the context of online platforms like Twitter, Facebook, and Reddit, and the affordances of these platforms can affect the way it is engaged with and perceived [5]. In particular, preliminary research has shown that user-generated comments can affect people's risk perception of the technologies discussed in science blogs [1], and even certain people's agreement of a science article's slant [70, 71], highlighting their potential importance in the way social media users interact with science news in the modern era.

Past research on the effects of user commentary on news perception has focused on typical blog and online news platform formats; that is, participants would read the text of an article, and then read the comments which follow [1, 28, 29, 69, 70]. However, social media platforms like Facebook and Reddit allow users to view comments before following the link to a posted article. In

Authors' address: Spencer Williams, sw1918@uw.edu; Gary Hsieh, garyhs@uw.edu, University of Washington, Seattle, WA, USA, 98195.

2573-0142/2021/4-ART32 $15.00
https://doi.org/10.1145/3449106

particular, recent research on Reddit's r/science sub-community suggested that its users often read user-generated comments before the article itself [22]. Because of this, comments may have the potential to signal various qualities of an article before a person even reads it, potentially impacting their likelihood of reading the article at all.

Furthermore, while past research has studied dimensions such as valence [15, 18, 66, 69], civility [1, 44], quality of reasoning [44], type of argument [70], and exemplification [56], comments in the science news domain may vary along other understudied dimensions. For example, given the well-known challenge of communicating scientific topics without using technical jargon [51], comments about science news may vary in terms of technical difficulty, which could yield different impressions about the article and its underlying research. In addition, comments that provide information about an article are among the most popular on r/science [22]; such comments may highlight that an article may contain useful information, or they may summarize the article well enough the users feel they no longer have to read it at all.

The present work examines how different types of user comments affect readers' perceptions of science articles posted in an r/science-like context, as well as their intention to read them at all. Study 1 examines both comments' agreement with the article and their difficulty, providing experimental evidence that difficult comments reduce people's likelihood to read the associated article, regardless of their particular slant. Study 2 is a pre-registered, confirmatory follow-up experiment that shows that people's expectations for science news articles are affected by a similarity heuristic; the attributes of a comment such as its difficulty, entertainment value, or information quality can signal those attributes of the article itself, impacting their decision to read the article or not. Ultimately, we show how different types of user commentary can impact people's interest in reading science news articles, and conclude with recommendations for the design of science communication communities.

2 RELATED WORK AND HYPOTHESES

2.1 Effects of User Commentary on News Perception

Over the past several years, social media has become an increasingly important source of science news [48]. The public has recently turned to mixed media for their science news, relying not only on traditional formats such as television and newspapers, but blogs and social media sites like Facebook, Twitter, and Reddit [58]. One important affordance of such online platforms is the presence of user comments [28].

While few studies to date have examined the interaction between user commentary and science news perception specifically, a significant body of work over the past decade has shown that user-generated comments in general can impact readers' perceptions of their associated articles in a number of ways [30]. For example, negatively-valenced comments have been shown to reduce readers' perceptions of article credibility [15, 66], issue importance, relevant behavioral intentions [56, 74] and the persuasive power of the article itself [15, 69–71], as well as activate prejudices and affect participants' perception of reality [29]. Comments that disagree with the article's message have also been shown to affect certain participants' personal opinions about the associated issue, as well as their inferences about public opinion [19, 28, 67].

Beyond the effects of specifically negative comments, others have examined how the quality of reasoning and level of civility in the comments section affects article perceptions. For example, the presence of poorly-reasoned comments has been shown to reduce people's expectations about the informational quality of the associated article [44]. The same study also showed that the presence of uncivil comments can decrease people's expectations about the "formal" quality (i.e. various factors associated with the quality of writing) of the article, even suggesting that the mere presence of any type of comment may reduce readers' quality expectations. Thus, it is possible that even positive or neutral comments could have a negative impact on people's perceptions of their associated articles.

Notably, the bulk of past research has shown user-generated comments to participants following the article rather than before [1, 19, 28, 29, 67, 69–71], mirroring the structure of typical online news outlets and blogs. However, while a few recent studies have noted this problem and studied how reading comments *before* an article can impact people's impressions of it [44, 66], these studies do not provide clear mechanisms for how different comment types can affect people's expectations for the article, nor do they investigate how those expectations might ultimately affect people's interest in reading the article itself. Given that social media sites like Facebook, Twitter, and Reddit present comments before users navigate to the posted articles, such effects could have significant implications for how people engage with the news on such sites, but current research on this topic is sparse.

Thus, the primary, overarching research question across both studies is concerned with filling in this existing gap in the literature, and understanding how different comments impact people's expectations of, and interest in reading, science news article:

RQ0: How do different types of user-generated comments affect people's expectations of related science news articles, and how do those expectations affect their interest in reading the articles themselves?

2.2 Science News on Social Media

Online blogging has become an important source of scientific and educational information for both expert and lay audiences [14], and individual researchers often make use of such channels in order to communicate their work to a wider audience [21]. Other social media channels, such as Twitter and Reddit, have also become important avenues for disseminating scientific information [22, 46, 65], and many researchers believe that communicating about their research through these channels can benefit the public overall [73].

In particular, Reddit may be especially relevant to the study of user commentary and science news, specifically with regard to its "sub-reddit" r/science. r/science is a hub for users to post and discuss recent science news, boasting over 22 million registered members [22]. Because of the way the site is structured, clicking on the title of a Reddit post immediately brings up the attached comments section, and past research has shown that r/science users often read the comments before the article itself [22], potentially coloring user's perception before they navigate to the original source. Given Reddit's importance as a news source for millions of people worldwide [39], this affordance could have significant impact on the way its user selectively engage with the news.

While a wide variety of comment types have been identified on r/science, one that regularly receives a high proportion of "upvotes" (and consequently, more visibility based on Reddit's algorithm for displaying comments) is the method critique [22]; comments that question the methods used in the underlying study may be among the top comments for a given post, and thus the first thing that many users would see after clicking the post title. Because of the numerous effects of negatively valenced comments for general news articles [15, 28, 66, 69–71], it is possible that these critical comments may also negatively impact reader's perceptions of the article's message, its persuasive power, or the credibility of the underlying science.

Beyond that, such critique may impact how likely a person is to read the associated article at all. In general, people show more interest in a task when they perceive it as having value or utility [16, 61]. Furthermore, people have been shown to be more likely to read news articles when they perceive them as having higher information utility [26]. If participants believe that the methods used in a study are suspect, and if they then decide that the study's findings are invalid, they may consequently perceive the article as having low information utility, which would in turn discourage them from reading the article, as there is less apparent worth in doing so. Thus, our first hypothesis for Study 1...

H1: People who read negative comments will show less interest in reading the associated article.

Of course, the above assumes that people are basing the usefulness of a task on its informational utility. It is possible that, for participants who typically engage with science news for other purposes (e.g. social, entertainment), they may still be interested in reading a paper whose methods have been criticized by others. Given this possibility, We hypothesize an interaction such that participants' motivations for news will impact what effects different user commentary will have:

H2: Participants who are motivated by the informational aspect of reading science news will be less likely to read an article with a negative comment, but that this effect will be weaker for participants who are motivated by factors other than information.

Finally, it is also unclear what effect positive critiques would have on readers' perceptions of the article. While the effects of negative comments are well-documented [15, 28, 66, 69, 70], previous research has consistently shown that positive comments do not have an equivalent effect [66, 69]. However, because positive research method critiques for science news have not yet been studied to our knowledge, it remains to be seen whether reading such opinions will increase people's agreement with the article's findings, or the quality of the underlying research.

RQ1: What effect, if any, will positive comments have on participants' interest in reading the article?

2.3 Difficulty of Scientific Texts

Balancing accurate scientific information while avoiding technical jargon is a well-known challenge for science writers [12, 51], and user-generated comments that are difficult to understand may be a barrier to broader accessibility in communities like r/science [22]. However, beyond creating a higher barrier to entry for lay audiences, the difficulty of a scientific text has been shown to impact people's judgements about its findings, their confidence in the material [48, 49], and even their interest in and identification with science as a whole [52]. For example, the presence of technical jargon [48] and arguments based on empirical evidence [25] may be more difficult for lay audiences to process, making them feel like they do not understand the article.

Thus, one potential effect of reading negative comments would be a decreased interest in reading the article. People who read difficult comment may become less interested in the article's topic, or they may feel like they will not understand the article. Thus, one hypothesis would be...

H3$_\alpha$: People who read a difficult comment will show less interest in reading the associated article.

However, it is possible that difficult comments may have the opposite effect. Recent research has described an "easiness effect" of science texts, where people who read more difficult articles show less confidence in their understanding of the material, whereas reading easier texts may make them overconfident [50]. In this case, people who read easier comments may feel like they already understand what the article is trying to say, and consequently feel less need to read it. In contrast, those who read a more difficult comment may have more interest in reading the article, in order to better understand the issue. Thus, as a competing hypothesis to $H3_\alpha$:

H3$_\beta$: People who read a difficult comment will show increased interest in reading the associated article.

In order to examine these hypotheses, we first present an experimental study to investigate the effects of user-generated comments on people's interest in reading science news articles.

3 STUDY 1: COMMENT VALENCE AND DIFFICULTY

Study 1 uses experimental methods to determine the effects of a comments' difficulty and agreement with the associated article on participants' interest in reading that article. By showing participants different versions of an r/science comment that varied along the above dimensions and surveying their intention to read it, we provide an initial exploration into how different types of comments impact people's interest in reading science news articles.

3.1 Methods

This study was a between subjects experiment with a 2 (comment's agreement with the article: positive or negative) × 2 (comment difficulty: easy or hard) + 1 (control) design, for a total of five conditions. Participants were shown a mockup of a reddit post (see Figure 1) that linked to a science news article, as well as one of the possible comments (or no comments in the control condition). In order to isolate the effects of specific comments, our study focused only on a single-comment scenario; however, while many comments sections on Reddit have few or no comments [72], our results may not fully generalize to posts with many comments, which we discuss in the limitations section. Once participants answered questions about the comment and their likelihood of reading the article, they were then directed to the article itself. The chosen article fit r/science's guidelines for appropriate material [45] based on quality, relevance, and recency from when our study was conducted, and was determined to be representative of a typical r/science article.

3.1.1 Participants. A total of 611 participants were recruited on Amazon Mechanical Turk (AMT). Because we wanted to sample from an average population of news readers, we chose AMT due to its demographics being more diverse than the typical college sample [9, 40] (gender-balanced, skewing slightly young, income comparable to US average) [9]. Additionally, online news reading behavior among AMT workers has been shown to reflect real-life behavior on social media [34], and studies of news perception and sharing behavior frequently sample from this population [6, 32, 59]. The study was advertised as "Read a science news article and answer a survey." After excluding those participants who failed the attention check, those who reported skipping the comment, those who could not answer what the comment was about, and those who dropped out of the survey partway through, the final sample size was 298 (age: M = 31.4, SD = 8.9). In terms of gender breakdown, 185 participants identified as male, 111 identified as female, and 2 identified as non-binary. The sample was 46.3% (138) White, 41.9% (125) Asian, 3.7% (11) Hispanic or Latino, 3.0% (9) Black or African American, 1.3% (4) American Indian or Alaska Native, 0.3% (1) Pacific Islander, and 1.0% (3) who chose not to say. 59% (176) of the sample reported using Reddit at least once a day, and 91% (271) reported using it at least once a week. Most (66.78%, 199) either "agreed" or "strongly agreed" that "Reddit is a good source of science news," with only 14.09% (42) who "disagreed" or "strongly disagreed."

3.1.2 Stimuli. The chosen article was a summary of a recent study done on the relationship between living near a major roadway and developmental delays in children (see Appendix A). Beyond being chosen for its appropriateness based on r/science's rules, it was also expected to be a topic of relative interest to the general public. The article was 671 words long, and was included within the questionnaire using a single, sans-serif font and with no attached advertisements.

All four comments focused on the authors' method of estimating air pollution in their study, given the high visibility of this type of comment on r/science [22]. The positive comments provided an argument for why "the authors' method of estimating air pollution was pretty clever," which focused on their use of workplace location data to better estimate pollution exposure throughout the day. The negative comments provided an argument for why "the authors' method of estimating air pollution was pretty poor," which was based on the lack of traffic data attached to the roadways being studied, which could have provided a more accurate estimate of air quality.

Difficulty was manipulated as per Scharrer, et al [50]. That is, the easy comments contained a direct causal explanation for their arguments, were shorter in length, and contained no complex language, abbreviations, or jargon. On the other hand, the difficult comments provided links to other studies and web pages to back up their claims (the empirical argument type described by [25]), contained abbreviations and other jargon, and were about twice as long. To illustrate, the positive and easy comment was written as follows:

Their method of estimating air pollution was pretty clever. Mothers whose workplaces were far from major roads would be exposed to less pollution, so taking that into account along with their home addresses allowed for a better estimate of exposure overall.

In contrast, the negative and hard comment was written this way (underlined text indicates a web link):

Their method of estimating air pollution was pretty poor. While they did control for the quantity of traffic, vehicle type was unaccounted for. Motor vehicle registration data shows significant differences in vehicle types throughout different states, and previous studies have shown a significant effect of vehicle type on CO, VOC, and PM emissions. Past studies have also shown that combining the International Vehicle Emissions (IVE) model with video road surveys provides an accurate breakdown of traffic-driven effects on air quality, which the authors failed to include in their own model.

In each condition, the headline and comment were displayed on a replica of an r/science page (See Figure 1). The post was set up to be 20 minutes old in order to foster agnosticism about the posts quality; it was meant to be recent enough that the cumulative upvotes would not be a useful indicator of public opinion about the article, while being old enough that at least one user could have feasibly read, digested, and commented on the article,

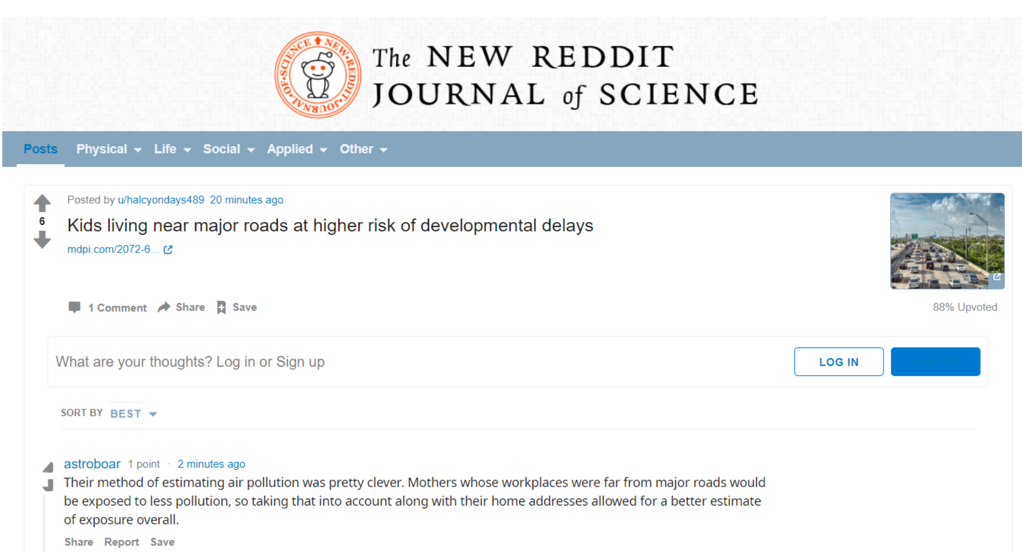

Fig. 1. r/science mockup used in Study 1. User names were generated specifically for this study and were not in-use by any actual Reddit users.

3.1.3 Measures. Participants' interest in reading the article were assessed on a 5-point Likert scale ("How likely would you be to read the article linked to in the previous screenshot?"). In order to ensure that participants actually found the technical jargon and empirical arguments to be more difficult to understand, we also included a manipulation check ("How easy or difficult was it to understand the comment in the previous screenshot?"), also on a 5-point scale.

All motivations indices were adapted from Park, Kee, and Valenzuela [41], with participants being asked to describe how strongly they agreed with the statements "I engage with science news (to feel connected to a community/to talk about research with others/to get useful information/to learn about topics I care about/because it's fun/because it's entertaining)." They were also asked how much they agreed with the statement, "I think Reddit is a good source of science news," as well as how often they visited Reddit and how often they read science news. Finally, all demographic questions were taken from Hughes, et al [20].

3.1.4 Design. At the beginning of the study, participants were instructed to "please go through the information just like you would normally if you were reading it on social media." The next page contained the r/science screenshot, where participants read through the headline and comment described above. In the experimental conditions, once they navigated to the next page, participants were then asked whether or not they had read the comment. Once they responded, they were asked about the difficulty of the comment and their interest in reading the article.

After completing those initial questions, participants then moved on to the article itself and were asked to read the article, regardless of their reported interest in reading the article. We designed our study this way because in addition to testing interest in reading, we also wanted to examine how likely participants were to share the associated article, before and after reading it. Specifically, we hypothesized that participants would be less likely to share articles when the comment was difficult, as it may reduce their confidence in the material. However, although the negative/hard comment reduced their interest in sharing the article before reading it ($p = .002$), there was no effect for the positive/hard comment ($p = .118$), nor was there an effect on sharing intention after reading the article for the negative/hard ($p = .131$) or positive/hard ($p = .588$) conditions. Given these null results, we focus our attention on interest in reading across both studies.

As per the above instructions, participants were allowed to spend as much time as they wanted reading it (seconds spent reading: $M = 57.8$, $SD = 90.2$). Upon completing the article, they were asked about their opinions of and habits using Reddit, their motivations for reading science news, and their demographic information.

3.2 Analysis

Analyses were primarily conducted under a Bayesian framework, as recommended for statistical analysis in HCI [24], although frequentist statistics are presented as well for reference. Analysis was done using the rstan package for R [62]. As per Winter, et al [69], we condense our $2 \times 2 + 1$ conditions into a 1×5 ANOVA, and examine pairwise differences to disentangle interaction effects. Our linear model was a modified version of the robust Bayesian ANOVA offered by Kruschke [27], with three modifications. First, while we retained the t-distribution for observations, we used a constant $\nu = 10$ rather than putting an exponential prior distribution on ν, which reduced divergent transitions during sampling. Second, we placed a uniform prior over σ, rather than a hierarchy of gamma distributions, also to reduce divergent transitions; the prior was placed over the range of 0-2 to allow for a wide range of realistic standard deviations. Third, we used a non-centered parameterization as described in the Stan User's Guide [62], which can improve sampling efficiency and accuracy while providing an otherwise identical model. Using the notation x[i] to indicate membership in one of the four comment conditions, our final model was defined as:

Table 1. Means and standard deviations of participants' interest in reading the article, by group.

	Interesting in reading
Negative/Hard	2.79 (1.39)
Negative/Easy	3.22 (1.26)
Positive/Hard	2.90 (1.45)
Positive/Easy	3.04 (1.43)
Control	3.49 (1.06)

$y = t(10, \mu, \sigma)$
$\mu = \beta[i] + \alpha$
$\sigma \sim uniform(0, 2)$
$\beta[i] = \beta_{raw}[i]\sigma_{\beta}[i]$
$\beta_{raw} \sim normal(0, 1)$
$\sigma_{\beta} \sim gamma(\alpha_{\gamma}, \beta_{\gamma})$
$\alpha \sim normal(M_y, SD_y)$

Where α_{γ} and β_{γ} are shape and rate parameters chosen such that the gamma distribution has a mode equal to half the standard deviation of y, and a standard deviation equal to twice the standard deviation of y, as recommended by Kruschke [27]. Sampling was done using four chains of 25,000 iterations each, with 2000 iteration warm-up periods. All reported parameter estimates had effective sample sizes greater than 30,000, r-hat values less than 1.01, and the models had no divergent transitions during sampling. In all cases, parameters whose 95% credible intervals did not overlap with 0 were considered significant, foregoing the use of Bayes factors based on the discussion by Kruschke [27]. These credible intervals are based on central intervals using Stan's defaults [62].

Means and standard deviations for participants' interest in reading the article are provided in Table 1. Plots of posterior distributions over β values (referred to here as the deflection parameters [27], which indicate the difference from the control group) for the four experimental conditions are given in Figure 2. Our Bayesian model was applied as described above, with four deflection parameters for each of the four experimental conditions. Because of the differences in how multiple comparisons are handled under frequentist versus Bayesian frameworks (see Gelman [13] for an in-depth discussion, or Kruschke [27] for applications to experimental research), our frequentist pairwise comparisons do not always match the Bayesian ones in significance. However, for both Study 1 and Study 2, we provide estimates from both frameworks to demonstrate robustness (or lack thereof) across multiple choices of analyses.

3.3 Results

3.3.1 Manipulation Check. In order to ensure that our manipulation of difficulty (technical jargon, empirical arguments) was appropriate, we ran our model on participants' ratings of comment difficulty, comparing the effects of difficult comments vs. easy comments. Participants in the difficult comment conditions (M = 2.57, SD = 1.13) rated the comment as being significantly more challenging to understand than participants in the easy comment conditions (M = 2.09, SD = 1.14), with a [0.21, 0.82] 95% credible interval over the difference between the two conditions. A frequentist t-test supports this finding, with t(235) = 3.29, p = .001, d = 1.68, indicating a very large effect [8]. Thus, the experimental manipulation was successful.

3.3.2 Main Results.

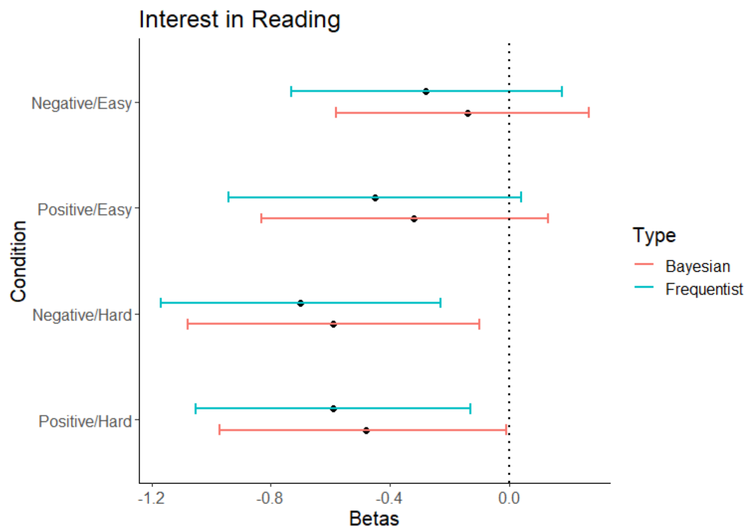

Fig. 2. Bayesian credible intervals and frequentist confidence intervals over the beta values describing the average change in likelihood to read for (from top to bottom) negative/easy, positive/easy, negative/hard, and positive/hard comments compared to control. The greater the distance between the CIs and 0, the greater confidence we have that the results for that condition are different than the control condition. Here, both difficult comments cause significantly decreased interest in reading compared to control.

Valence. We found that H1, predicting that negative valenced comment will decrease people's interest in reading, is not supported. Specifically, our analyses showed that while the negative/hard comment showed significant deflection from control, β = -0.58 [-1.08, -0.10], the negative/easy comment was not significant, β = -0.14 [-0.58, 0.27]. We also fail to support H2, which predicted that participants who were motivated by the information utility of science news would show reduced interest in reading the article following a negative comment. Motivation scores were calculated using the statistical average of the inventory items. The interaction (information × valence) was not significant (p = .193).

Difficulty. Finally, we support H3$_\alpha$ (and reject H3$_\beta$), which predicted that difficult comments would cause decreased interest in reading the article. The positive/hard comment showed significant deflection from control β = -0.48 [-0.97, -0.01], as did the negative/hard comment as described above. This is mirrored by a significant frequentist ANOVA, F(4, 293) = 2.68, p = .032, η_p^2 = 0.035, indicating a small effect [8]. Furthermore, in terms of RQ1, we do not find evidence that the positive conditions improved participants' interest in reading the article. To ensure that participants' previous exposure to science news was not responsible for this result, we also ran an ANCOVA using how often participants read science news as a control variable, F(4, 293) = 2.91, p = .022. Results mirrored the above, with the negative/hard (p = .004) and positive/hard (p = .025) conditions remaining significant under this model, suggesting our results were not caused by this individual difference.

Interestingly, while only the difficult comments yielded significantly reduced interest in reading compared to the control, all four comment conditions trended in the negative direction, and pooling all comments into a single predictor in our Bayesian model shows that seeing any of the four comments, on average, decreased participants' interest in reading the article, β = -0.46 [-0.84, -0.07]. We will address this general trend in the discussion section.

3.4 Discussion

Study 1 provides a novel examination of how user-generated comments can affect people's interest in reading the associated article; interestingly, and in contrast with our expectations, our findings suggest that reading both positive and negative comments—specifically difficult comments—may reduce people's willingness to read the article, rather than the main effect of negative comments that we predicted. While past research has noted multiple negative effects of difficult scientific text [48, 49, 52], and we indeed show that the presence of difficult comments can reduce people's interest in reading the article, it is not clear precisely what accounts for this effect in this context.

One explanation may be that reading a difficult comment could signal that the article itself is more difficult, which could explain why both difficult article conditions showed significantly reduced likelihood to read; given that people tend to avoid texts they expect will be more difficult [55], a comment that provides this signal could deter them from reading. It may also be that they feel less connection to science in general after reading a difficult comment [52], which could also contribute to this effect.

We also note that difficult comments did not increase people's interest in reading the article, despite the potential consequences of the easiness effect [50]. Perhaps even if participants did feel less confident about their understanding of the study prior to reading it, they may not have felt like reading the article would have been helpful in understanding it (e.g. if the above explanation is correct and they assumed the article would be just as difficult to understand). In that case, there would be little utility in reading the article, as reading an equally difficult text may do little to improve their understanding.

Ultimately, the results of Study 1 show a general trend for our difficult comments to reduce participants likelihood to read. However, the reason for this effect remains unclear. Thus, Study 2 is an attempt to glean insight about the potential reasons reading comments might affect how likely people are to read the article.

4 STUDY 2: DIFFICULTY, INFORMATION QUALITY, AND AROUSAL

Study 2 provides a deeper look into the different types of comments that can impact people's interest in reading science articles, and what mediating variables may be responsible for this effect. While there are a number of potential explanations for the effects of difficult comments to reduce people's interest in reading the associated article in Study 1, including increased expectations of the article's difficulty or reductions in its expected information quality, we lacked the data to explore these possibilities. Thus, our second experiment is a pre-registered, confirmatory follow-up testing these possible explanations and clarifying the effects of user comments on people's perceptions of science articles.

4.1 Hypotheses

In Study 1, we found evidence that reading a difficult comment before a science news article can impact people's interest in reading that article, but were unable to conclude what may have been driving this effect. In Study 2, we will both extend and deepen our understanding of the ways in which different comments can impact people's interest in reading science articles, by examining three distinct features: difficulty, information quality, and entertainment. By examining comments that vary along these three dimensions, and exploring their effects on a number of potential mediating variables, we can better determine what may have accounted for the effects we observed in Study 1.

4.1.1 Difficulty. First, in terms of difficulty, we expect that our results from Study 1 will replicate. However, there are a few possible mechanisms for this effect. First, past research has found that the presence of comments can impact people's expectations of an article's quality [44], based on certain heuristic cues [66]. Perhaps when viewing user-generated comments, people utilize a similarity heuristic [60], whereby certain information (in this case, a comment) is being used as a signal for the type of information to expect (in this case, the article) [42]. If this were the case, it may be that if people read a difficult comment, they will expect that the article itself will be difficult to understand, thus reducing their interest in reading it. Thus:

H4a$_\alpha$: People who read a difficult comment will expect the article to be more difficult to read than those who read less difficult comments.

However, given that recent work has found that the use of jargon can reduce people's self-reported connection to the scientific community, as well as their interest in science at all, these constructs may also come into play when people read more technical comments. Further, if jargon reduces people's interest in science, they may no longer consider scientific articles as personally relevant, due either to feelings of alienation from science [63], difficulty finding personally useful information [42], or the assumption that more technical articles are not written for public consumption. This in turn could reduce their interest in reading the article if they are motivated by personal interests or affirmation [37]:

H4a$_\beta$: People who read more difficult comments will show less connection to the scientific community, less interest in science in general, and less interest in the specific article than people who read easier comments.

In either case, we expect that people who read more difficult comments will show less interest in the article than people who read easier comments.

H4b: Participants who read more difficult comments will show less interest in reading the article.

4.1.2 Information Quality. As noted above, prior work [44] has suggested that user comments may signal something about the quality of an article. Given this, it may be that after reading a comment that is poorly-written or unhelpful, one may assume that the article it is associated with will also be low-quality. While previous research has examined "quality" both in terms of "formal quality" (i.e. how well an article was written) as well as "informational quality" (i.e. how useful and accurate the information in an article will be), we expect that this latter quality will be more closely associated with people's likelihood to read an article, given the importance of information motivation in selecting online news [37]. Thus, we expect that if a comment contains useful information, it will signal to readers that the article will be similarly useful. That is:

H5a: Participants who read comments with low-information quality will expect the article to contain less useful information than participants in other conditions.

H5b: Participants who read comments with low-information quality will be less likely to read the article.

4.1.3 Entertainment. Finally, one of the most common motivations for reading online news is entertainment [37]. Because of this, we expect that if reading an entertaining comment provides a signal that the article itself may be more entertaining, we would expect that people who read entertaining comments may be more interested in reading their associated article. Thus:

H6a: Participants who read more entertaining comments will expect that the article will be more entertaining as well.

H6b: Participants who read the more entertaining comment will be more likely to read the article.

H6c: There will be an interaction such that participants who are more entertainment-motivated will be affected by entertaining comments than participants who are not.

In addition, we explore a number of other variables here, such as participants' agreement with the study's methods, agreement with the findings, participants' confidence in their judgements, their interest in the article's subject, and their perception of the commenter's expertise. Although we do not provide any specific hypotheses here or in our preregistration (and thus these analyses should be considered exploratory), we include these here to explore whether factors other than a comment's agreement with the article might affect people's decisions about the methods or findings of a study, hoping to tentatively build on past research [70].

4.2 Methods

Study 2 utilized a 3 (easy vs. low information vs. entertaining) + 1 (Base comment) + 1 (Control) experimental design. Rather than using a fully crossed design as in Study 1 (which would have required a total of 9 conditions, and a prohibitively large sample size), we instead designed a base comment meant to reflect a typical r/science comment [22]. This allowed us to compare each of the experimental conditions with a comment that varied along one of the experimental dimensions, as well as a no-comment control as in Study 1. It was preregistered [1] in order to explicitly lay out our confirmatory hypotheses to differentiate between exploratory results [36], guard against p-hacking [54], and help to promote more open research and support better meta analyses in HCI [7].

4.2.1 Participants. We recruited a total of 423 participants on Amazon Mechanical Turk. We again included multiple attention checks, and sampled participants until we retained a total of 286 (age: M = 35.0, SD = 10.57). 192 of our participants identified as male, 90 identified as female, 1 identified as non-binary, and 3 chose not to specify. The sample was 66.5% (176) White, 21.0% (60) Asian, 5.6% (16) Black or African American, 4.9% (14) Hispanic or Latino, 0.4% (1) American Indian or Alaska Native, and 0.4% (1) Native Hawaiian or Pacific Islander. 54.5% (156) reported using Reddit at least once a day, and 82.5% (236) using at at least once a week. These numbers were notably lower for r/science usage, with only 11.5% (36) visiting the subreddit once a day, and 38.5% (110) once a week.

4.2.2 Stimuli. As per Study 1, we chose a science news article and created an r/science mockup of it. This new article discussed the relationship between having a more optimistic romantic partner and long-term cognitive health (see Appendix B). Comments varied based on how much information they contained, how much technical jargon they contained, and the tone of the writing (see Table 2). The base comment was designed to mimic typical popular comments on r/science [22], and contained useful information about the study, technical jargon, and was written in a non-arousing manner. The easy comment was identical to the base comment, but replaced the technical jargon with plain-language explanations of the concepts. In contrast, the low-quality information comment contained a positive evaluation of the study, but no other information. Finally, the entertaining comment contained more exclamations (to increase arousal, which is related to entertainment [38]) and jokes.

In every other way, the design of the r/science mockup was identical to Study 1, with the exception of the control condition. In order to control for the possibility that participants might use the lack of comments itself as a signal about the quality of an article, we shifted the location of the screenshot to remove the comments section entirely, introducing ambiguity about whether any comments existed (see Figure 3)

4.2.3 Measures. All previous measures from Study 1 were retained, and all new measures were 5-point Likert scales, except where specified. We measured information quality for both the comment and the article using an index adapted from Bobkowski, et al. [3] (e.g. "How (helpful/valuable/useful/informative) do you think the information in the article will be?"). We also adapted questions about people's interest in science ("How interested are you in scientific topics in general?") and people's identification with science ("I am good at science," "It is important to

[1]Link to anonymized registration: https://osf.io/pq9v5?view_only=a57d43b09ab14f938d6c18d9fe36d613

Table 2. Descriptions of comment types.

Comment	Difficulty	Information	Entertainment	Text
Base	High	High	Low	This study was actually pretty strong, and the evidence seems to suggest a clear link between having an optimistic partner and long-term health. By using a large sample (N = 8,914; M age = 66.73, SD = 9.67) from the longitudinal HRS database over 8 years (t1; t2; t3; t4; t5), as well as multilevel dyadic analysis to account for non-independence in their data, they find significant associations between partner optimism and actor cognitive functions like memory (r = .04, p < .05) and mental status (r = .03, p < .05) while controlling for possible confounds. Overall, it appears to be a small but clear effect.
Easy	Low	High	Low	This study was actually pretty strong, and the evidence seems to suggest a clear link between having an optimistic partner and long-term health. By using a large sample with nearly 9,000 participants from the Health and Retirement Study database over 8 years, as well as using advanced statistical techniques, they find significant associations between a person's partner's optimism and that person's cognitive functions like memory and mental status, while controlling for other economic and health factors. Overall, it appears to be a small but clear effect.
Low-quality	Low[2]	Low	Low	This study was pretty good.
Entertaining	High	High	High	This study was actually really cool! It's a convincing case for a link between having an optimistic partner and long-term health (makes me think I should put "glass half full people only" on my Tinder profile...). By using a large sample (N = 8,914; M age = 66.73, SD = 9.67) from the longitudinal HRS database over 8 years (t1; t2; t3; t4; t5), as well as multilevel dyadic analysis to account for non-independence in their data, they find significant associations between partner optimism and actor cognitive functions like memory (r = .04, p < .05) and mental status (r = .03, p < .05) while controlling for possible confounds. Overall, it appears to be a small but clear effect.

me that I am good at science") from Shulman, et al. [52]. Additional variables included people's interest in the article's subject ("How interested are you in the article's topic?"), the importance of the article's subject ("How important do you think the article's topic is?"), the expected difficulty of

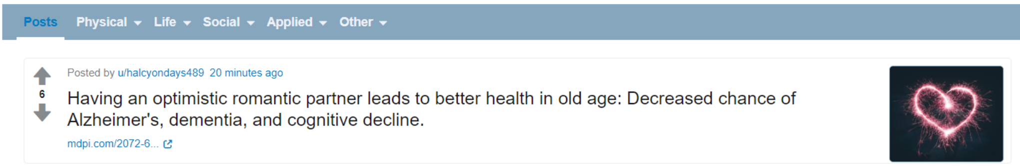

Fig. 3. r/science mockup used in Study 2.

the article ("How difficult do you think it would be to understand the article?"), and the expected entertainment value of the article ("How entertaining do you think the article will be?"). We also added manipulation checks for comment information quality and comment entertainment value. These were displayed immediately before participants read the article, as in Study 1.

Additionally, we included a set of variables about participants' impressions of the article after having read it. These included their agreement with the study's methods ("I think the researchers used appropriate methods in the study described above."), its findings ("I think having an optimistic romantic partner leads to better health in old age."), and their confidence in their judgement of the article ("I am confident that my answer to the above question is correct.").

We also added a set of open-ended questions for participants to fill out, after the post-article DVs. These asked for their expectations about the article ("Thinking back to when you first saw the Reddit snapshot, how did the information on the page (the headline, any comments you saw, etc.), affect what you thought about the article?") and their likelihood to read the article specifically ("How, if at all, did that affect your likelihood to want to read the article?").

In every other way, the design of Study 2 was identical to Study 1.

4.3 Analysis

Our Bayesian model was identical to Study 1, except that in order to reduce divergent transitions, we tuned our model to have a slightly tighter prior over σ_β by multiplying the rate parameter of the gamma distribution by 1.5, one of several strategies recommended for dealing with divergences [33]. Sampling was done with 4 chains of 15,000 iterations each, with 2000 iteration warm-up periods. All reported parameter estimates had effective sample sizes greater than 20,000, r-hat values less than 1.01, and no divergent transitions. As in Study 1, we provide frequentist tests for reference.

Our pre-registered analyses included using our Bayesian model to analyze each of our DVs and mediators, combined with frequentist ANOVAs. Because the dimensions of our experimental manipulation (difficulty, information quality, entertainment) were not fully crossed, we utilize a modified Dunnett's test [4] for our frequentist models to compare each experimental condition to both the base comment condition and the no-comment control condition, without comparing the experimental condition to each other. Because we do not apply family-wise error rate corrections to the Bayesian models, we report multiple comparison results as-is between the experimental and control conditions, but again we do not report comparisons between the experimental conditions.

We also provide exploratory mediation [64] and path analyses [47]. Because we did not specify models for these in our pre-registration, these should not be considered confirmatory. Rather, we discuss these results along with our qualitative data to help clarify the experimental effects we

[2]Because of its lack of content, the low-quality comment was also easy to understand. Mediation analysis ensures that our experimental results for this condition were driven by differences in information quality rather than difficulty.

Table 3. Means and standard deviations by group. Superscripts b and c mean the condition is significantly difference from the base and control conditions (p < .05) respectively.

DVs	Read	Agree	Confidence	Methods		
Entertaining	2.73 (1.15)	4.20 (0.72)	4.30 (0.66)	3.95 (0.75)		
Easy	3.62 (1.16)^b	4.43 (0.54)	4.22 (0.99)	4.05 (0.80))		
Low-quality	2.91 (1.20)	4.07 (0.76)^b	3.96 (0.98)^b	3.68 (0.76)^b		
Base	2.95 (1.22)	4.49 (0.57)	4.42 (0.65)	4.31 (0.70))		
Control	3.22 (1.29)	4.17 (1.02)	4.14 (0.99)	3.95 (0.94)		
Mediators	Difficulty	Information	Entertainment	Sci Interest	Sub Interest	Sci Identity
Entertaining	2.38 (1.20)	3.70 (0.94)	2.54 (1.13)^c	3.63 (0.91)	2.95 (1.23)	3.01 (1.01)
Easy	2.15 (1.29)	4.20 (0.56)^c	2.89 (1.18)	3.89 (0.90)	3.44 (1.03)	3.05 (1.10)
Low-quality	2.02 (1.22)^b	3.56 (0.84)^b	2.70 (1.19)^c	3.84 (0.80)	3.11 (1.19)	3.20 (0.98)
Base	2.73 (1.11)^c	4.08 (0.63)	2.64 (1.24)^c	3.54 (1.06)	3.07 (1.13)	2.96 (1.22)
Control	2.08 (1.28)	3.79 (0.88)	3.31 (0.95)	3.78 (0.95)	3.46 (1.01)	3.19 (1.03)

observed, and follow-up studies may be needed to confirm the effects of some of these factors on people's likelihood to read science articles.

4.4 Results

4.4.1 Manipulation Checks. First, we find that our difficulty manipulation was successful. Specifically, there were significant deflections for both the easy comment β = -1.73 [-2.08, -1.37], as expected, and the low-quality comment β = -1.70 [-2.06, -1.35] (likely due to its short length and lack of information). These results were echoed by a significant ANOVA, F(3, 223) = 46.72, p < .001, η_p^2 = .386), suggesting a very strong effect [8].

Our information quality manipulation was also successful, with significant effects for both the low-quality comment, β = -1.10 [-1.44, -0.77], as well as the entertaining comment, β = -0.46 [-0.78, -0.14], supported by a highly significant anova, F(3, 223) = 22.60, p < 0.001, η_p^2 = .285, indicating a large effect [8]. Although the manipulation for the low-quality comment was successful, it may be that the presence of humor in the entertaining comment reduced participants' impressions of the information provided in that comment as well, which may impact how the entertaining comment affected their impressions of the article itself beyond the intended manipulation.

Moving on to entertainment, we did not detect a significant difference between the entertaining comment and the base comment, β = -0.01 [-0.40, 0.37] (frequentist μ_Δ = -0.03, p = .999), suggesting that this manipulation was not successful. However, there were significant differences between the easy and base comments for entertainment value, β = 0.48 [0.04, 0.93] (marginally significant under the frequentist model μ_Δ = 0.48, p = .077), suggesting that the easy comment was perceived as more entertaining than the base (i.e. more difficult) comment.

4.4.2 Main hypothesis tests.

Difficulty. See Table 3 for means and standard deviations. Starting with difficulty (see Figure 4), our results support H4a$_\alpha$. We find that the base (i.e. difficult) comment increased expected article difficulty compared to the easy comment, μ_Δ = -0.62 [-1.06, -0.18]. The difficult comment also showed significant deflection from control (β = 0.64 [0.21, 1.06]), with an overall significant frequentist anova (F(4, 281) = 4.89, p = .011) and a marginally significant difference between the easy and base conditions (μ_Δ = -0.58, p = .062).

Moving on, our results only partially support H4a$_\beta$. Participants who read the more difficult (base) comment showed a decreased interest in science compared to the easy comment, μ_Δ = 0.34

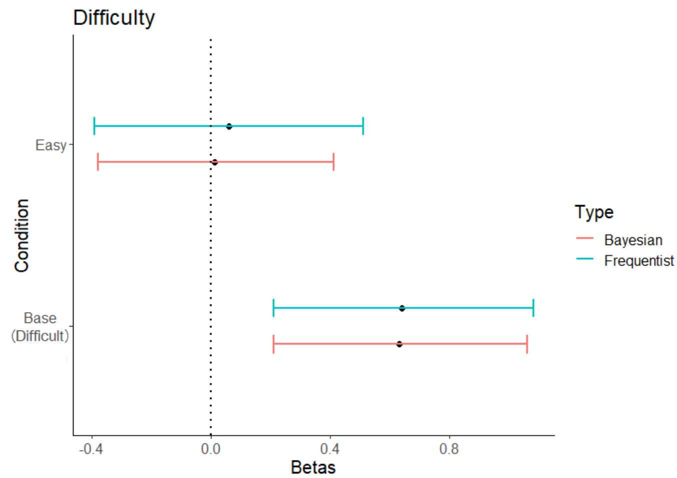

Fig. 4. Bayesian and frequentist CIs over effects of comment types on expected article difficulty.

[0.02, 0.67]. However, the frequentist ANOVA was not significant, F(4, 281) = 1.45, p = .217, η_p^2 = .020, suggesting that this effect was not strong enough to show robust significance given our sample size. Moreover, there were no effects on participants' science identity via either framework, with a non-significant frequentist ANOVA F(4, 281) = 0.59, p = .668, η_p^2 = .008. There were also no significant pairwise effects on their interest in the article's topic via either framework, but the ANOVA was significant in this case, F(4, 281) = 2.42, p = .049, η_p^2 = .033. Overall, because of the disagreement between the Bayesian and frequentist frameworks, and the partial support for an effect on science and subject interest but not on science identity, we consider H4a$_\beta$ to be only somewhat supported.

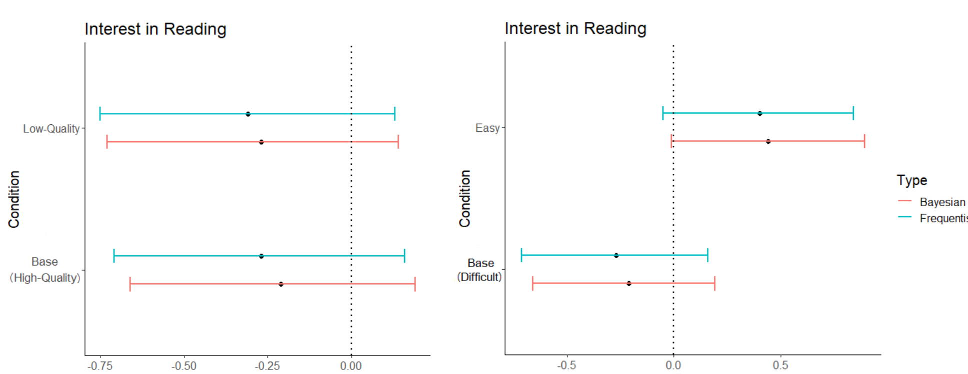

Fig. 5. Bayesian and frequentist CIs over effects of comment types on interest in reading the article.

Finally, we look at participants' interesting in reading the article (see Figure 5), with results supporting H4b. Participants in the easy condition showed significantly more interest in reading the article than participants in the base (difficult) condition under both the Bayesian (μ_Δ = 0.65 [0.20, 1.11]) and frequentist (μ_Δ = 0.64, p = .026) frameworks (although neither were significantly different from control under either framework), with a significant overall ANOVA F(4, 281) = 3.31, p = .011, η_p^2 = .061 indicating an effect size somewhat larger than Study 1 (although still relatively small [8]).

Information Quality. Moving on, our results support H5a; we found that the low-quality comment significantly reduced participants' expectations of the article's information quality (see Figure 6), μ_Δ = -0.44 [-0.72, -0.17], although the difference between the low-quality and control condition did not quite reach significance β = -0.25 [-0.52, 0.02]. The frequentist ANOVA was also significant, F(4, 281) = 6.54, p < .001, η_p^2 = .085, with a significant difference between low-quality and base conditions (μ_Δ = -0.52, p = .003). However, we fail to support H5b; there were no significant differences between the low-quality comment and either comparison condition on participants' interest in reading the article (see Figure 5).

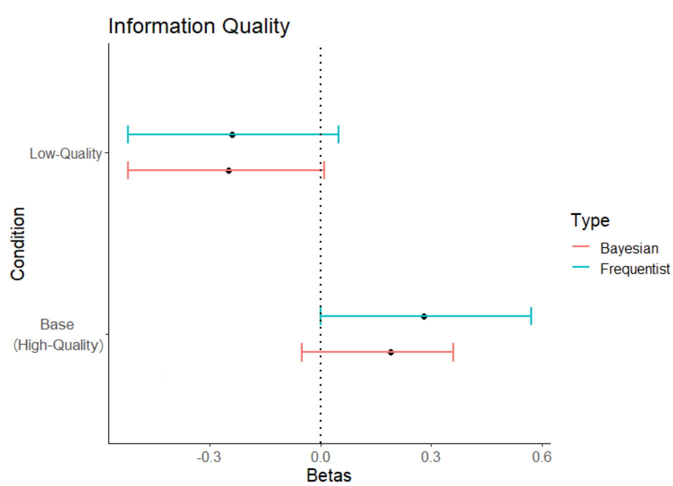

Fig. 6. Bayesian and frequentist CIs over effects of comment types on expected article information quality.

Because the low-quality comment was also significantly easier to understand than the base comment, we use mediation analysis to ensure that our experimental results were accounted for by the difference in information quality, rather than any possible effect of difficulty (see Section 4.4.4 for more details on our mediation analysis methods). Indeed, we show that the effect of the low-quality comment on expected article quality was significantly mediated by the comment's quality (average causal mediated effect (ACME) = -0.62, p < .001).

Entertainment value. Because the manipulation check for the entertaining comment was unsuccessful, we use participants' ratings of the comments' entertainment value as a predictor for their expected article entertainment value instead to test H5a (see Table 4). Indeed, we show that participants' expectations for the article's entertainment value was significantly predicted by their assessment of the comment's entertainment value, supporting H6a. However, because this ad-hoc analysis was not specified in the pre-registration, these tests should not be considered confirmatory.

Moving on to H6b, we also find that comment entertainment significantly predicted participants' interest in reading the article (see Table 5), tentatively supporting this hypothesis (although this main effect loses significance under the interaction model, see below). We also hypothesized that entertainment-motivated participants would be more motivated to read articles after reading high-arousal comments. Again, because the entertaining condition failed our manipulation check, we

Table 4. Summary of linear model on participants' expected article entertainment value as a function of their perceived comment entertainment.

Expected Article Entertainment			
Predictors	Estimate	CI [95%]	p
(Intercept)	1.18	[0.93, 1.42]	< .001
Comment Entertainment	0.68	[0.58, 0.78]	< .001
Observations	286		
Adjusted R^2	=.468		
p	< .001		

Table 5. Summary of linear model on participants' interest in reading as a function of their perceived comment entertainment.

Interest in Reading			
Predictors	Estimate	CI [95%]	p
(Intercept)	1.81	[1.52, 2.10]	< .001
Comment Entertainment	0.55	[0.44, 0.67]	< .001
Observations	286		
Adjusted R^2	=.292		
p	< .001		

examine the interaction term between entertainment motivation and comment entertainment value; however, there was no significant interaction (p = .461), so H6c was also not supported.

4.4.3 *Exploratory variables.*

Agreement with methods. Moving on to our more exploratory analyses (see Table 3 for descriptive statistics), we first examine participants' agreement with the methods of the study. Unexpectedly, and in contrast with past research [69], we found that the base comment increased participants' agreement with the study's methods compared to control (β = 0.30 [0.02, 0.57]), whereas the low-quality comment actually decreased participants' agreement with the methods compared to control (β = -0.29 [-0.57, -0.01]), with a strong difference between the two conditions (μ_Δ = -0.59 [-0.87, -0.30]). This is supported by a significant ANOVA, F(4, 281) = 4.60, p = .001, η_p^2 = .061, and although the differences between control and base (μ_Δ = 0.36, p = .085) and low-quality (μ_Δ -0.26, p = .318) lose significance under the frequentist framework, there is still a highly significant difference between the low-quality and base comments (μ_Δ = -0.62, p < .001).

Agreement with findings. In terms of participants' agreement with the findings themselves, although none of the deflections from control were significant, there was still a similar difference between the low-quality and base comments, with participants who read the low-quality comments showing less agreement with the study's results (μ_Δ = -0.36 [-0.62, -0.11]). This finding holds up under a significant frequentist ANOVA, F(4, 281) = 3.39, p = .010, η_p^2 = .046, echoing the same pairwise difference (μ_Δ = -0.42, p = .016).

Confidence. Furthermore, the pattern of results for participants' confidence in their judgements was similar, with no significant deflections from control but a significant difference between the low-quality and base comments (μ_Δ = -0.36 [-0.66, -0.06]), with participants who read the base comment showing more confidence. Although the overall frequentist ANOVA was only marginally significant in this case, F(4, 281) = 2.229, p = .060, η_p^2 = .032, there was still a significant pairwise difference under the frequentist framework (μ_Δ = -0.46, p = .029).

4.4.4 Mediation tests. For our mediation tests, we employ Tingley, et al.'s [64] mediation package for R. This allows us to estimate the *average causal mediation effects* (ACME), and the *average direct effects* (ADE) for each level of the "treatment" variable (in this case, different comment types). A significant ACME suggests that a significant amount of the effect of some variable X on outcome variable Y is mediated by some variable M. A significant ADE, on the other suggests, that there is a significant effect of X on Y after controlling for M. Because we did not pre-specify our models in our pre-registration, and because we made ad-hoc adjustments to our analyses in order to better disentangle our effects, all proceeding analyses should be considered exploratory. All analyses are bootstrapped simulations using 10,000 iterations each. Each mediation test is done using one of the experimental conditions as a predictor, comparing against the base and control comments as a baseline.

First, we investigate whether the change in expected article difficulty mediated by the positive effect of reading easier comments (see Table 6). However, there was no significant mediated effect, although the overall difference between the easy comment and the baseline comments was significant. We also examined the potential mediating effect of expected entertainment value, but unsurprisingly, there was also no significant mediating effect, given that even the easy condition failed to significantly increase this variable. However, because of the significant effect of the easy comment on expected information quality, we examined information quality as a mediator as well, and found that it did significantly mediate the positive effect of the easy comment, with the easy comment leading to increased expected information quality, and consequently, higher interest in reading the article. We will examine this effect further in the proceeding section.

We also examined the difference between the low-quality comment and the baseline comments, due to its strong effect on information quality. Similar to the easy comment, we found that the low-quality comment's effect was also significantly mediated by expected information quality, although in the opposite direction; reading the low-quality comment reduced people's expectations about the article's information quality, which reduced their interest in reading the article, although as the non-significant total effect and experimental results show, the overall effect was not strong enough to show significance in this case.

While these analyses provide some insight into the effects of different comment types on participants' interest in reading the article, the relationships between our full set of mediators, and how they lead participants to read or avoid associated articles, is still unclear. We next turn to structural equation modeling to better clarify a possible structure whereby our experimental manipulation and observed mediators impacted participants' self-reported likelihood to read the article.

4.4.5 Structural Equation Modeling. We began by examining a model relating the comment variables (difficulty, information quality, and entertainment) with the article variables (information quality, difficulty, entertainment, and topic interest), which in turn predicted interest in reading. We included participants' overall interest in the article's topic as a predictor, allowing us to control for this variable in the regressions and improve model fit. After dropping variables which showed no relationship with any outcome variables and making ad-hoc adjustments to ensure good model fit, we fit the final model (see Figure 7) as follows:

read ~ difficulty$_a$ + entertainment$_a$ + info$_a$ + topic interest
info$_a$ ~ difficulty$_a$ + info$_c$ + topic interest
entertainment$_a$ ~ entertainment$_c$ + topic interest
difficulty$_a$ ~ difficulty$_c$

Table 6. Summary of our mediation analyses.

Difficulty (Comment)→Difficulty (Article)→Read			
Quantity	Value	CI [95%]	p
ACME	-0.04	[-0.14, 0.02]	= .246
ADE	0.57	[0.19, 0.95]	= .004
Total Effect	0.53	[0.15, 0.92]	= .006
Proportion Mediated	-0.06	[-0.49, 0.05]	= .250
Information (Comment)→Information (Article)→Read			
Quantity	Value	CI [95%]	p
ACME	-0.23	[-0.43, -0.07]	= .004
ADE	0.06	[-0.31, 0.44]	= .755
Total Effect	-0.17	[-0.57, 0.23]	= .391
Proportion Mediated	0.87	[-10.17, 12.26]	= .392
Difficulty (Comment)→Information (Article) →Read			
Quantity	Value	CI [95%]	p
ACME	0.18	[0.04, 0.35]	= .013
ADE	0.35	[-0.02, 0.72]	= .065
Total Effect	0.53	[0.14, 0.92]	= .007
Proportion Mediated	0.34	[0.08, 1.06]	= .018
Difficulty(Comment) →Entertainment (Article) →Read			
Quantity	Value	CI [95%]	p
ACME	-0.04	[-0.20, 0.12]	= .667
ADE	0.57	[0.21, 0.92]	= .002
Total Effect	0.53	[0.14, 0.93]	= .008
Proportion Mediated	-0.06	[-0.83, 0.22]	= .675

where the subscript a indicates the expected value of that variable for the article, and the subscript c indicates the perceived value of the comment. While the model fit was acceptable (CFI = .963, SRMR = .069, RMSEA = .080), another important criterion for evaluating SEMs is their theoretical interpretability [31], so we will briefly discuss our modeling choices and interpretation as we discuss results.

Overall, because our experimental results showed that participants who read comments of a certain attribute (e.g. more technically difficult, more useful information) showed higher expectations that the article itself would have those attributes (e.g. participants who read the easy comment expected the article to be easier to understand), we modeled those article mediators as being predicted by their comment-based counterparts. Furthermore, it stands that the expected entertainment and informational value of an article will both depend on how interested one personally is in the topic, a relationship which we see here.

If we consider the overall rate of information gain as a ratio of the value of the information over the cost of retrieving it [42], we would expect that information quality would be negatively impacted by expected article difficulty; if an article is more difficult to understand, it may not be considered as useful, since the information within it becomes harder to parse. Indeed, we show a negative relationship between information quality and difficulty in our model.

We also expected that the four primary article mediators would all be associated with participants' interest in reading it. However, in this case, only entertainment and information quality were

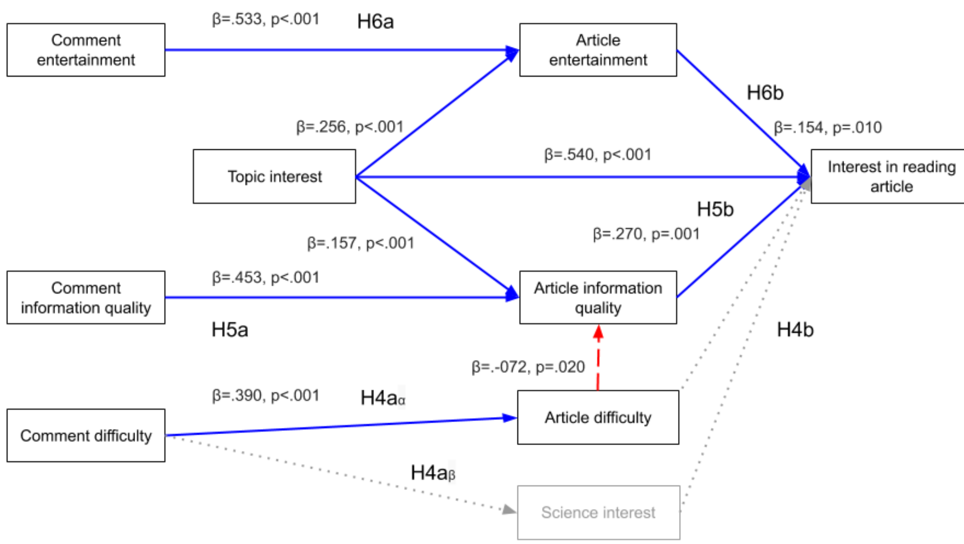

Fig. 7. Path diagram describing our model of mediating variables on participants' interest in reading the science article. Solid blue lines indicate a positive relationship, whereas dashed red lines indicate a negative relationship. Overall, we observe that comment attributes signal that the article will share similar attributes, each of which may somehow affect their final decision to read that article. We include unsupported hypothesized paths (dotted gray lines) for illustrative purposes, although they were not included in the final SEM specification due to lack of experimental support and negative effects on model fit.

significant, although difficulty may still have an indirect relationship via its reduction on expected information quality.. However, topic interest itself did show a strong relationship with interest in reading, somewhat unsurprisingly. Overall, we show that the attributes of a comment seem to be related to participants' expectations for the article, many of which may be linked to their likelihood to read it.

4.4.6 Open-Ended Results. For each comment condition, the first author coded participants' answers to the open-ended questions as to whether they mentioned the comment, and if so, whether it increased or decreased their likelihood to read the article. Overall, we found that participants drew a number of inferences about the article from the different comment types, largely supporting our quantitative findings. Quotes are lightly edited for grammar and spelling.

First, in our quantitative analyses, we found that difficult comments may be used heuristically to infer article difficulty. In our qualitative responses, participants mentioned how the more technical and difficult comments can affect their perceptions of the article and subsequent likelihood to read it:

> The comment was too scientific and too wordy...It made me think that the article had too many scientific terms and that it would be confusing and hard to understand, so the comment made me not want to read the article initially.
>
> I did not want to read the article if it was filled with terms that I could not easily understand. Having to research the terms would have been too much of a hassle for me to read about this topic.

Similarly to comment difficulty, we also found that comment's information quality affect perceived information quality of the article.

> Having only one short comment without any information about it made me think poorly about the article...The poor comment probably would make me skip reading the article.

> The comment was super descriptive and informative and made me think that there was a lot of worthy information to be found in the article...Sometimes I read the comment sections of an article first to see what other people think of it, so yes a very well written informative comment would for sure effect the likelihood of myself reading an article.

However, participants also reported some other relationships that we did not explore and deserve future explorations. First, there seemed to be an effect of summarizing too much information, with some participants mentioning that they felt like the comment had enough information that they would no longer be interested in reading the article itself:

> I felt like I got a nice summary of the article in the comment and felt it not necessary to have to read the full article so I would have never clicked on the link myself or have shared it.

Second, although not found in our model, participants did report topic interest as a potential moderator.

> If the topic was something that I was very passionate about then having to do extra research would not be a problem.

4.5 Discussion

Study 2 provided a confirmatory follow-up to Study 1. We confirmed that the presence of technical jargon in a comment can reduce people's interest in reading the attached article compared to people who read an easier comment. Seeking to understand the mechanism at work, we found that reading difficult comments can increase the expected difficulty of the article. However, our path analyses suggest that difficulty affects people's likelihood indirectly through article information quality. Perceived article difficulty by itself did not influence likelihood to read; it negatively influenced people's perceptions on the usefulness of the information in the article, which then affected their likelihood to read the article.

We also showed that the quality of a comment can impact the expected quality of the article. This may provide insight on previous research which found that the presence of poorly-reasoned and uncivil comments reduced participants' impressions of the information and writing quality, respectively, of a news article [44]. That is, if a poorly-reasoned comment is considered to have low-quality information, readers may use that as a signal to judge the quality of the article itself.

Furthermore, unlike some previous research [69], we observed a strengthening effect whereby reading a positive, technical comment actually increased participants' agreement with the study's methods, while the low-quality comment reduced it. This may be due to the strong effect our comment conditions had on participants' confidence; perhaps the presence of complex jargon signaled that the commenter had more expertise, or that their arguments were more sound, providing readers with more confidence in their arguments. In contrast, because we showed that user comments can signal the quality of the article, poor-quality comments could reduce users' expectations about the article, which may include its methods and findings. Thus, even a positive comment has the potential to reduce people's expectation of an article, if it is otherwise poorly written or unhelpful.

Finally, although our results showed that a comment with high-quality information will reflect well on the article itself, some of our participants described how reading a comment that summarizes the findings of an article may decrease their likelihood to read it, believing that they have already learned most of its relevant information. Given these observations, it may be that reading high-quality comments will improve people's expectations of the article's quality, but that summarizing the article too fully may actually have a negative effect, reducing their interest in reading the article. Thus, it is important to understand what types of comments are more likely to elicit one effect or the other.

For example, certain types of comments may serve as better summaries of the article than others, a possibility which may help explain the general negative trend for all comments in Study 1. That is, in Study 1, the article was based on a phenomenon (kids living near major roadways suffer developmental delays) which was both fairly intuitive (living by the highway causes exposure to harmful pollutants) and unlikely to affect most people. Reading a comment that explains whether or not the study was trustworthy may be enough for most people, who can easily extrapolate how the study might affect them personally. On the other hand, Study 2 used two constructs (trait optimism and long-term cognitive health) that may be less intuitive (e.g. what counts as "optimism?"), and it may not be as clear how the one affects the other. Where participants in Study 1 may have felt like the comment was enough for them in most cases, this may have been the case less often in Study 2, due to participants requiring more information on average. In either case, further research may be needed to determine how the potential conflicting effects of information quality and summaries in comments may interact.

5 GENERAL DISCUSSION

Across two experiments, we identify numerous effects of user-generated comments on people's expectations of science news articles. We show that reading user comments before an article can signal the difficulty, information quality, and entertainment value of an article. Specifically, we find that people utilize a similarity heuristic; they infer that the article will hold similar characteristics as the article's comments (e.g. difficulty). While some previous research [1, 70] has examined the effects of user commentary in a science news context, they were done using a typical blog format with comments presented after the article. Here, we show that when people read the comments first (as many do on sites like Reddit [22]), it can color their perception of the article in numerous ways.

Our studies also help elucidate the impact these perceptions of the article may have on people's likelihood to read the articles. Through our path analyses, we found that perceived article entertainment value and information quality both directly increase people's interest in reading the article. Article difficulty, however, did not have a direct effect. Rather, it decreased interest in reading by reducing the expected information quality of the article.

Having this similarity heuristic towards online news is understandable. Because of the vast amount of information online, people are going to rely on heuristics to help them determine how to allocate their attention [60]. In a comment-first design, people will utilize the available information (i.e., the comments) to help them make their inferences. However, problems arise when the qualities of the comments do not match the qualities of the articles. This could lead to people reading bad science because others have provided high quality comments, or more problematically, high quality articles being overlooked because of low-quality comments. Furthermore, as we have found, the low-quality comments may not just influence likelihood to read, but also how strongly people agree with the study's methods or findings. This also means that campaigns to discredit science can be extremely effective in the comment-first design. Low quality comments may not only turn people away from the articles, but also make them think poorly of the studies themselves.

5.1 Design Implications

Our findings suggest a number of design implications for science news communities. Had we found evidence that an "easiness effect" [50] in user comments would discourage people who read easy comments from seeking out more information, it may have useful to highlight such comments on communities like r/science. However, given the overall negative effect of difficult comments in Study 1 and the difference between easy and difficult comments in Study 2, we suggest the opposite; these communities should explicitly discourage overly-technical comments, which could prevent them from signalling that an article is more difficult, and thus less informative, than it may actually be. Currently, r/science has no guidelines on the use of technical language for comments [45], and this could have the added benefit of ensuring that discussions are accessible to non-expert audiences as well, although balancing accessibility with scientific rigor is a well-known challenge [12, 22, 51]. Another solution could be for sites to automatically detect difficult comments and reduce their likelihood of being displayed up-front. This could be done via automated methods, either through sophisticated natural-language processing, or basic keyword-matching to identify technical jargon based on a pre-defined dictionary [53]. Of course, such a system should ideally be made explicit to the community and combined with written guidelines, to ensure that users' content is not suppressed in an opaque manner without their understanding.

Of course, we also found that the expected information quality of the article was the most direct mediator for people's interest in reading. So, because we found that the quality of the comments can signal the quality of the article itself, ensuring the comments people see before reading an article are themselves high-quality may be important. As it is, many social media sites provide these comments up front, which may encourage people to read them first [22], so ensuring those comments reflect well on the article is key. However, while r/science's moderation combined with Reddit's up-vote system typically ensure that high-quality comments are the most visible, other science communication communities on platforms like Facebook may not self-curate quality content as effectively [2]. As our study highlights, such communities should likely prioritize placing high-quality discussion up-front, and making such priorities explicit to the community.

However, our qualitative results suggest that some people may use the comments as a summary of the article itself, making them feel like they don't need to read the article at all. To side-step this potential issue, science communication platforms could consider ways of signaling these attributes without requiring users to read the comments themselves. For example, rather than a simple up-vote/down-vote system, platforms could implement more nuanced feedback about an article's quality, such as its technical difficulty, scope, validity, and depth of information, which could be available at a glance without needing to dive into the comments section. By allowing users to "up-vote" an article along one of the dimensions we found that could affect their interest in reading the article , more useful information about the article could bubble up, allowing users to make informed decisions about whether an article is worth reading, given their background, interests, and motivations.

Furthermore, by surfacing this information, platforms may be able to avoid potential spiral-of-silence [35] dilemmas. The spiral-of-silence, whereby people's assumptions of popular opinion prevent them from discussing contentions topics, can be a significant problem for science communication [43]. If users read a low-quality comment or difficult comment, they may decide not to read the article itself, coming away with an incomplete or false impression of the article [17]. This could be especially problematic given the tendency for people to assume that comments reflect public opinion [66]; an inaccurate comment may dissuade people from reading an article, while simultaneously providing an inaccurate impression of public opinion, potentially dissuading people from discussing a topic or voicing alternative opinions [35]. Through such a mechanism, user comments may have a significant impact on people's final impressions of a research finding [70], and low-quality or poorly-reason comments [44] that promote negative attitudes toward science (e.g. climate change denial [2]) may dissuade people from reading science articles and impact their opinions of relevant topics. Thus, providing mechanisms to reduce the negative effects of certain user comments and encouraging people to read science articles will help ensure such misunderstandings are minimized, and improve science communication on social media platforms like Reddit.

5.2 Future Research

Finally, while we have examined a number of different comment types and mediating variables, there are many potential directions for future work. For example, because of the potential conflicting effects of high–quality comments increasing the expected quality of the article, while also providing enough information that some people choose not to read the article itself, future research could focus on disentangling these effects. Perhaps comments that provide information about methods and study validity can increase expected information gain, while comments that discuss the nuance of the findings and potential applications may decrease it, as they may provide all the practical knowledge that a casual reader may need.

Furthermore, participants in our experiments were only presented with one comment before being asked to read the study. While this has allowed us to isolate the effects that one comment may have on participants' expectations for an article, further research could examine how these effects extend to busier comment sections. For example, if people read a mix of high- and low-quality comments, what assumptions will they make about the article? If two otherwise high-quality comments disagree with each other about the study's methods or findings, what strategies are used to determine who is correct, and consequently, whether a study is worth reading? Given that mixed frames can have different effects compared to the same messages in isolation [10], how do the properties of an entire comment section, and users' interactions with it, ultimately affect their likelihood to engage with the article itself?

More work may also be needed to generalize this study. First, although many of our participants had extensive experience with Reddit, it is unclear how well this effect might generalize to other online environments, such as Twitter or Facebook, where science news may also be shared. Second, although previous work shows that many readers spend more time on the comments than the article itself [57], and while many r/science users may read the comments first [22], it remains unclear how many people do this across platforms. So, surveys on how often people read comments before science news articles may be useful in understanding how frequently the effects we have identified might appear in practice.

Similarly, future work can examine whether this effect generalizes beyond user comments to other science communication domains. For example, citizen science has become a useful way to connect the broader public with research, allowing data collection and analysis at massive scales [68]. Recently, arguments have been made that researchers have an ethical obligation to share study information with online participants [23], treating this as another channel for science communication. While providing descriptions of online studies (e.g. in recruitment blurbs) could be a useful way of contextualizing them, they may also signal attributes of the study or citizen science endeavor (e.g. a technical blurb could make a task seem more difficult). Thus, future work could examine how online study participants or citizen scientists use limited information to judge research opportunities, and determine whether a similarity heuristic may be at play.

6 LIMITATIONS

While we set out to examine the effects of certain types of user comments on people's engagement with science news, this is only one possible context where comments might affect people's expectations about news articles. While we show that reading difficult comments can decrease people's interest in reading science news, it is unclear whether this effect would generalize to other domains, such as politics, business, art, etc. Furthermore, it is unclear how these effects would interact with the presence of other comments. Does reading a difficult comment still affect the expected difficulty of an article in the presence of easier comments? If there are a mix of high- and low-quality comments, how do people judge the quality of the article itself? Further experimentation and field studies may be required to better determine the generalizability and robustness of the effects we have uncovered.

7 CONCLUSION

Through multiple studies, we sought to better understand the effects of user-generated comments on people's engagement with and perception of science news on social media. Using an r/science context, we show that comments can impact people's expectations of an article's difficulty, quality, informativeness, and entertainment value, reduce or increase their agreement with the study's methods and findings, and in some cases, even their interest in reading the article itself. We show that participants use a similarity heuristic when making inferences about the quality of a science news article, which they then use to determine whether or not to read it at all. By exploring these effects, we hope to better inform science communication practices on social media, as well as support the missions of science-based communities like r/science to disseminate quality research.

REFERENCES

[1] Ashley A Anderson, Dominique Brossard, Dietram A Scheufele, Michael A Xenos, and Peter Ladwig. 2014. The "Nasty Effect:" Online Incivility and Risk Perceptions of Emerging Technologies. *Journal of Computer-Mediated Communication* 19, 3 (2014), 373–387. https://doi.org/10.1111/jcc4.12009

[2] Emma Frances Bloomfield and Denise Tillery. 2019. The Circulation of Climate Change Denial Online: Rhetorical and Networking Strategies on Facebook. *Environmental Communication* 13, 1 (2019), 23–34. https://doi.org/10.1080/17524032.2018.1527378

[3] Piotr S Bobkowski. 2015. Sharing the news: Effects of informational utility and opinion leadership on online news sharing. *Journalism & Mass Communication Quarterly* 92, 2 (2015), 320–345. https://doi.org/10.1177/1077699015573194

[4] Frank Bretz, Torsten Hothorn, and Peter Westfall. 2016. *Multiple comparisons using R*. CRC Press. https://doi.org/10.1201/9781420010909

[5] Dominique Brossard. 2013. New Media Landscapes and the Science Information Consumer. *Proceedings of the National Academy of Sciences* 110, Supplement 3 (2013), 14096–14101. https://doi.org/10.1073/pnas.1212744110

[6] Myojung Chung. 2017. Not Just Numbers: The Role of Social Media Metrics in Online News Evaluations. *Computers in Human Behavior* 75 (2017), 949–957. https://doi.org/10.1016/j.chb.2017.06.022

[7] Andy Cockburn, Carl Gutwin, and Alan Dix. 2018. Hark no More: On the Preregistration of CHI Experiments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 141. https://doi.org/10.1145/3173574.3173715

[8] Jacob Cohen. 1977. *Statistical Power Analysis for the Behavioral Sciences*. Academic press. https://doi.org/10.1016/c2013-0-10517-x

[9] Djellel Difallah, Elena Filatova, and Panos Ipeirotis. 2018. Demographics and Dynamics of Mechanical Turk Workers. In *Proceedings of the eleventh acm international conference on web search and data mining*. ACM, 135–143. https://doi.org/10.1145/3159652.3159661

[10] James N Druckman. 2010. Competing Frames in a Political Campaign. *Winning with words: The origins and impact of political framing* (2010), 101–120.

[11] Sharon Dunwoody. 2014. Science Journalism: Prospects in the Digital Age. In *Routledge handbook of public communication of science and technology*. Routledge, 43–55. https://doi.org/10.4324/9780203483794

[12] Elaine Howard Ecklund, Sarah A James, and Anne E Lincoln. 2012. How Academic Biologists and Physicists View Science Outreach. *PloS one* 7, 5 (2012), e36240. https://doi.org/10.1371/journal.pone.0036240

[13] Andrew Gelman, Jennifer Hill, and Masanao Yajima. 2012. Why We (Usually) Don't Have to Worry About Multiple Comparisons. *Journal of Research on Educational Effectiveness* 5, 2 (2012), 189–211. https://doi.org/10.1080/19345747.2011.618213

[14] Alison J Head, Michele Van Hoeck, and Kirsten Hostetler. 2017. Why Blogs Endure: A Study of Recent College Graduates and Motivations for Blog Readership. *First Monday* 22, 10 (2017). https://doi.org/10.5210/fm.v22i10.8065

[15] Dominique Heinbach, Marc Ziegele, and Oliver Quiring. 2018. Sleeper Effect from Below: Long-Term Effects of Source Credibility and User Comments on the Persuasiveness of News Articles. *New media & society* 20, 12 (2018), 4765–4786. https://doi.org/10.1177/1461444818784472

[16] Suzanne Hidi and K Ann Renninger. 2006. The Four-Phase Model of Interest Development. *Educational psychologist* 41, 2 (2006), 111–127. https://doi.org/10.1207/s15326985ep4102_4

[17] Laura Hlavach and William H Freivogel. 2011. Ethical Implications of Anonymous Comments Posted to Online News Stories. *Journal of Mass Media Ethics* 26, 1 (2011), 21–37. https://doi.org/10.1080/08900523.2011.525190

[18] Seoyeon Hong and Glen T Cameron. 2018. Will Comments Change Your Opinion? The Persuasion Effects of Online Comments and Heuristic Cues in Crisis Communication. *Journal of Contingencies and Crisis Management* 26, 1 (2018), 173–182. https://doi.org/doi.org/10.1111/1468-5973.12215

[19] J Brian Houston, Glenn J Hansen, and Gwendelyn S Nisbett. 2011. Influence of User Comments on Perceptions of Media Bias and Third-Person Effect in Online News. *Electronic News* 5, 2 (2011), 79–92.

[20] Jennifer L Hughes, Abigail A Camden, and Tenzin Yangchen. 2016. Rethinking and Updating Demographic Questions: Guidance to Improve Descriptions of Research Samples. *Psi Chi Journal of Psychological Research* 21, 3 (2016), 138–151. https://doi.org/10.24839/2164-8204.jn21.3.138

[21] Paige Brown Jarreau and Lance Porter. 2018. Science in the Social Media Age: Profiles of Science Blog Readers. *Journalism & Mass Communication Quarterly* 95, 1 (2018), 142–168. https://doi.org/10.1177/1077699016685558

[22] Ridley Jones, Lucas Colusso, Katharina Reinecke, and Gary Hsieh. 2019. r/science: Challenges and Opportunities in Online Science Communication. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 153. https://doi.org/10.1145/3290605.3300383

[23] Eunice Jun, Blue A Jo, Nigini Oliveira, and Katharina Reinecke. 2018. Digestif: Promoting Science Communication in Online Experiments. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–26. https://doi.org/10.1145/3274353

[24] Matthew Kay, Gregory L Nelson, and Eric B Hekler. 2016. Researcher-Centered Design of Statistics: Why Bayesian Statistics Better Fit the Culture and Incentives of HCI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 4521–4532. https://doi.org/10.1145/2858036.2858465

[25] Frank C Keil. 2010. The Feasibility of Folk Science. *Cognitive science* 34, 5 (2010), 826–862. https://doi.org/10.1111/j.1551-6709.2010.01108.x

[26] Hyun Suk Kim, Heather Forquer, Joseph Rusko, Robert C Hornik, and Joseph N Cappella. 2016. Selective Exposure to Health Information: The Role of Headline Features in the Choice of Health Newsletter Articles. *Media psychology* 19, 4 (2016), 614–637. https://doi.org/10.1080/15213269.2015.1090907

[27] John Kruschke. 2014. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press.

[28] Eun-Ju Lee and Yoon Jae Jang. 2010. What do Others' Reactions to News on Internet Portal Sites Tell Us? Effects of Presentation Format and Readers' Need for Cognition on Reality Perception. *Communication research* 37, 6 (2010), 825–846. https://doi.org/10.1177/0093650210376189

[29] Eun-Ju Lee, Hyun Suk Kim, and Jaeho Cho. 2017. How User Comments Affect News Processing and Reality Perception: Activation and Refutation of Regional Prejudice. *Communication Monographs* 84, 1 (2017), 75–93. https://doi.org/10.1080/03637751.2016.1231334

[30] Eun-Ju Lee and Edson C Tandoc Jr. 2017. When News Meets the Audience: How Audience Feedback Online Affects News Production and Consumption. *Human Communication Research* 43, 4 (2017), 436–449. https://doi.org/10.1111/hcre.12123

[31] Pui-Wa Lei and Qiong Wu. 2007. Introduction to Structural Equation Modeling: Issues and Practical Considerations. *Educational Measurement: Issues and Practice* 26, 3 (2007), 33–43. https://doi.org/10.1111/j.1745-3992.2007.00099.x

[32] Scott R Maier, Paul Slovic, and Marcus Mayorga. 2017. Reader Reaction to News of Mass Suffering: Assessing the Influence of Story Form and Emotional Response. *Journalism* 18, 8 (2017), 1011–1029. https://doi.org/10.1177/1464884916663597

[33] Martin Modràk. 2018. Taming Divergences in Stan Models. https://www.martinmodrak.cz/2018/02/19/taming-divergences-in-stan-models/

[34] Mohsen Mosleh, Gordon Pennycook, and David G Rand. 2020. Self-Reported Willingness to Share Political News Articles in Online Surveys Correlates with Actual Sharing on Twitter. *Plos one* 15, 2 (2020), e0228882. https://doi.org/10.31234/osf.io/zebp9

[35] Elisabeth Noelle-Neumann. 1974. The Spiral of Silence a Theory of Public Opinion. *Journal of communication* 24, 2 (1974), 43–51. https://doi.org/10.1111/j.1460-2466.1974.tb00367.x

[36] Brian A Nosek, Charles R Ebersole, Alexander C DeHaven, and David T Mellor. 2018. The Preregistration Revolution. *Proceedings of the National Academy of Sciences* 115, 11 (2018), 2600–2606. https://doi.org/10.31219/osf.io/2dxu5

[37] Heather O'Brien, Luanne Freund, and Stina Westman. 2013. What Motivates the Online News Browser? News Item Selection in a Social Information Seeking Scenario. *Information Research* 19 (2013).

[38] Mary Beth Oliver and Anne Bartsch. 2010. Appreciation as Audience Response: Exploring Entertainment Gratifications beyond Hedonism. *Human Communication Research* 36, 1 (2010), 53–81. https://doi.org/10.1111/j.1468-2958.2009.01368.x

[39] Steven Ovadia. 2015. More than Just Cat Pictures: Reddit as a Curated News Source. *Behavioral & Social Sciences Librarian* 34, 1 (2015), 37–40. https://doi.org/10.1080/01639269.2015.996491

[40] Gabriele Paolacci and Jesse Chandler. 2014. Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science* 23, 3 (2014), 184–188. https://doi.org/10.1177/0963721414531598

[41] Namsu Park, Kerk F Kee, and Sebastián Valenzuela. 2009. Being Immersed in Social Networking Environment: Facebook Groups, Uses and Gratifications, and Social Outcomes. *CyberPsychology & Behavior* 12, 6 (2009), 729–733. https://doi.org/10.1089/cpb.2009.0003

[42] Peter Pirolli and Stuart Card. 1999. Information Foraging. *Psychological Review* 106, 4 (1999), 643. https://doi.org/10.1037/0033-295X.106.4.643

[43] Susanna Hornig Priest. 2006. Public Discourse and Scientific Controversy: A Spiral-of-Silence Analysis of Biotechnology Opinion in the United States. *Science Communication* 28, 2 (2006), 195–215. https://doi.org/10.1177/1075547006293918

[44] Fabian Prochazka, Patrick Weber, and Wolfgang Schweiger. 2018. Effects of Civility and Reasoning in User Comments on Perceived Journalistic Quality. *Journalism Studies* 19, 1 (2018), 62–78. https://doi.org/10.1080/1461670x.2016.1161497

[45] Reddit. 2019. r/science. https://www.reddit.com/r/science/

[46] Julie M Robillard, Thomas W Johnson, Craig Hennessey, B Lynn Beattie, and Judy Illes. 2013. Aging 2.0: Health Information About Dementia on Twitter. *PLoS One* 8, 7 (2013), e69861. https://doi.org/10.1371/journal.pone.0069861

[47] Yves Rosseel. 2012. Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of statistical software* 48, 2 (2012), 1–36.

[48] Mike S Schäfer. 2017. How Changing Media Structures are Affecting Science News Coverage. *The Oxford Handbook of the Science of Science Communication* (2017), 51–57. https://doi.org/10.1093/oxfordhb/9780190497620.013.5

[49] Lisa Scharrer, Rainer Bromme, M Anne Britt, and Marc Stadtler. 2012. The Seduction of Easiness: How Science Depictions Influence Laypeople's Reliance on Their own Evaluation of Scientific Information. *Learning and Instruction* 22, 3 (2012), 231–243. https://doi.org/10.1016/j.learninstruc.2011.11.004

[50] Lisa Scharrer, Yvonne Rupieper, Marc Stadtler, and Rainer Bromme. 2017. When Science Becomes too Easy: Science Popularization Inclines Laypeople to Underrate their Dependence on Experts. *Public Understanding of Science* 26, 8 (2017), 1003–1018. https://doi.org/10.1177/0963662516680311

[51] Aviv J Sharon and Ayelet Baram-Tsabari. 2014. Measuring Mumbo Jumbo: A Preliminary Quantification of the use of Jargon in Science Communication. *Public Understanding of Science* 23, 5 (2014), 528–546. https://doi.org/10.1177/0963662512469916

[52] Hillary C Shulman, Graham N Dixon, Olivia M Bullock, and Daniel Colón Amill. 2020. The Effects of Jargon on Processing Fluency, Self-Perceptions, and Scientific Engagement. *Journal of Language and Social Psychology* (2020), 0261927X20902177. https://doi.org/10.1177/0261927x20902177

[53] Advaith Siddharthan. 2014. A Survey of Research on Text Simplification. *ITL-International Journal of Applied Linguistics* 165, 2 (2014), 259–298. https://doi.org/doi.org/10.1075/itl.165.2.06sid

[54] Joseph P Simmons, Leif D Nelson, and Uri Simonsohn. 2011. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological science* 22, 11 (2011), 1359–1366. https://doi.org/10.1037/e519702015-014

[55] M Cecil Smith. 2000. The Real-World Reading Practices of Adults. *Journal of Literacy Research* 32, 1 (2000), 25–52. https://doi.org/10.1080/10862960009548063

[56] Patric R Spence, Kenneth Lachlan, Timothy Sellnow, Robert G Rice, and Henry Seeger. 2017. That is so Gross and I have to Post About it: Exemplification Effects and User Comments on a News Story. *Southern Communication Journal* 82, 1 (2017), 27–37. https://doi.org/10.1080/1041794x.2016.1265578

[57] Natalie Jomini Stroud, Emily Van Duyn, and Cynthia Peacock. 2016. Survey of Commenters and Comment Readers. *Report, The University of Texas at Austin, US* 14 (2016).

[58] Leona Yi-Fan Su, Heather Akin, Dominique Brossard, Dietram A Scheufele, and Michael A Xenos. 2015. Science News Consumption Patterns and their Implications for Public Understanding of Science. *Journalism & Mass Communication Quarterly* 92, 3 (2015), 597–616. https://doi.org/10.1177/1077699015586415

[59] Emily Sullivan, Dimitrios Bountouridis, Jaron Harambam, Shabnam Najafian, Felicia Loecherbach, Mykola Makhortykh, Domokos Kelen, Daricia Wilkinson, David Graus, and Nava Tintarev. 2019. Reading News with a Purpose: Explaining User Profiles for Self-Actualization. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and*

Personalization. 241–245. https://doi.org/10.1145/3314183.3323456

[60] S Shyam Sundar. 2008. The MAIN model: A Heuristic Approach to Understanding Technology Effects on Credibility. *Digital media, youth, and credibility* 73100 (2008). https://doi.org/10.1162/dmal.9780262562324.073

[61] Ayumi Tanaka and Kou Murayama. 2014. Within-Person Analyses of Situational Interest and Boredom: Interactions between Task-Specific Perceptions and Achievement Goals. *Journal of Educational Psychology* 106, 4 (2014), 1122. https://doi.org/10.1037/a0036659

[62] Stan Development Team. 2019. Stan User's Guide. https://mc-stan.org/docs/2_18/stan-users-guide/

[63] Jessica J Thompson and Mark Windschitl. 2005. "Failing Girls": Understanding Connections among Identity Negotiation, Personal Relevance, and Engagement in Science Learning from Underachieving Girls. *Journal of Women and Minorities in Science and Engineering* 11, 1 (2005). https://doi.org/10.1615/jwomenminorscieneng.v11.i1.10

[64] Dustin Tingley, Teppei Yamamoto, Kentaro Hirose, Luke Keele, and Kosuke Imai. 2014. Mediation: R Package for Causal Mediation Analysis. (2014). https://doi.org/10.18637/jss.v059.i05

[65] Giuseppe A Veltri and Dimitrinka Atanasova. 2017. Climate Change on Twitter: Content, Media Ecology and Information Sharing Behaviour. *Public Understanding of Science* 26, 6 (2017), 721–737. https://doi.org/10.1177/0963662515613702

[66] T Franklin Waddell. 2018. The Authentic (and Angry) Audience: How Comment Authenticity and Sentiment Impact News Evaluation. *Digital Journalism* (2018), 1–18. https://doi.org/10.1080/21670811.2018.1490656

[67] Joseph B Walther, David DeAndrea, Jinsuk Kim, and James C Anthony. 2010. The Influence of Online Comments on Perceptions of Antimarijuana Public Service Announcements on YouTube. *Human Communication Research* 36, 4 (2010), 469–492. https://doi.org/10.1111/j.1468-2958.2010.01384.x

[68] Andrea Wiggins and Kevin Crowston. 2011. From Conservation to Crowdsourcing: A Typology of Citizen Science. In *2011 44th Hawaii international conference on system sciences*. IEEE, 1–10. https://doi.org/10.1109/hicss.2011.207

[69] Stephan Winter, Caroline Brückner, and Nicole C Krämer. 2015. They Came, They Liked, They Commented: Social Influence on Facebook News Channels. *Cyberpsychology, Behavior, and Social Networking* 18, 8 (2015), 431–436. https://doi.org/10.1089/cyber.2015.0005

[70] Stephan Winter and Nicole C Krämer. 2016. Who's Right: The Author or the Audience? Effects of User Comments and Ratings on the Perception of Online Science Articles. https://doi.org/10.1515/commun-2016-0008

[71] Stephan Winter, Nicole C Krämer, and Yuhua Jake Liang. 2017. User-Generated Opinion: How Reader Reactions and Source Reputation Influence the Effects of Online News. *SCM Studies in Communication and Media* 6, 3 (2017), 240–261. https://doi.org/10.5771/2192-4007-2017-3-240

[72] Max Woolf. 2014. *A Statistical Analysis of 142 Million Reddit Submissions*. Retrieved May 21, 2020 from https://minimaxir.com/2014/12/reddit-statistics/

[73] Yimei Zhu and Kingsley Purdam. 2017. Social Media, Science Communication and the Academic Super User in the United Kingdom. (2017). https://doi.org/10.5210/fm.v22i11.7866

[74] Marc Ziegele, Christina Koehler, and Mathias Weber. 2018. Socially Destructive? Effects of Negative and Hateful User Comments on Readers' Donation Behavior toward Refugees and Homeless Persons. *Journal of Broadcasting & Electronic Media* 62, 4 (2018), 636–653. https://doi.org/10.1080/08838151.2018.1532430

Received June 2020; revised October 2020; accepted December 2020