

Forthcoming at *Nature Human Behaviour*

Negativity drives online news consumption

Claire E. Robertson^{1,†}, Nicolas Pröllochs^{2,†}, Kaoru Schwarzenegger³, Philip Pärnamets⁴, Jay J.

Van Bavel^{1,5} & Stefan Feuerriegel⁶

¹ Department of Psychology, New York University, New York, NY, U.S. 10003, ORCID: 0000-0001-8403-6358

² Department of Business and Economics, University of Giessen, 35394 Giessen, Germany, ORCID: 0000-0002-1835-7302

³ Department of Management, Technology and Economics, ETH Zurich, 8092 Zurich, Switzerland

⁴ Division of Psychology, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden, ORCID: 0000-0001-8360-9097

⁵ Center for Neural Science, New York University, New York, NY, U.S. 10003, ORCID: 0000-0002-2520-0442

⁶ LMU Munich School of Management, LMU Munich, 80539 Munich, Germany, ORCID: 0000-0001-7856-8729

† These authors contributed equally

Correspondence. Jay J. Van Bavel (jay.vanbavel@nyu.edu) and Stefan Feuerriegel (feuerriegel@lmu.de).

Abstract

Online media is important for society in informing and shaping opinions, hence raising the question of what drives online news consumption. Here, we analyze the causal effect of negative and emotional words on news consumption using a large online dataset of viral news stories. Specifically, we conducted our analyses using a series of randomized controlled trials ($N = 22,743$). Our dataset comprises ~105,000 different variations of news stories from Upworthy.com that generated ~5.7 million clicks across more than 370 million overall impressions. Although positive words were slightly more prevalent than negative words, we found that negative words in news headlines increased consumption rates (and positive words decreased consumption rates). For a headline of average length, each additional negative word increased the click-through rate by 2.3%. Our results contribute to a better understanding of why users engage with online media.

1 Introduction

The newsroom phrase “if it bleeds, it leads” was coined to reflect the intuition among journalists that stories about crime, bloodshed, and tragedy sell more newspapers than stories about good news¹. However a large portion of news readership now occurs online—the motivation to sell papers transformed into a motivation to keep readers clicking on new articles. In the United States, 89% of adults get at least some of their news online, and reliance on the Internet as a news source is increasing². Even so, most users spend less than 5 minutes per month on all of the top 25 news sites put together³. Hence, online media is forced to compete for the extremely limited resource of reader attention⁴.

With the advent of the Internet, online media has become a widespread source of information and, subsequently, opinion formation^{5,6,7,8,9}. As such, online media has a profound impact on society across domains such as marketing^{10,11}, finance^{12,13,14}, health¹⁵, and politics^{16,17,18,19}. Therefore, it is crucial to understand exactly what drives online news consumption. Previous work has posited that competition pushes news sources to publish “click-bait” news stories, often categorized by outrageous, upsetting, and negative headlines^{20,21,22}. Here, we analyze the effect of negative words on news consumption using a massive online dataset of viral news stories from Upworthy.com—a website that was one of the most successful pioneers of click-bait in the history of the Internet²³.

The tendency for individuals to attend to negative news reflects something foundational about human cognition—that humans preferentially attend to negative stimuli across many domains^{24,25}. Attentional biases towards negative stimuli begin in infancy²⁶ and persist into adulthood as a fast and automatic response²⁷. Furthermore, negative information may be more

“sticky” in our brains; people weigh negative information more heavily than positive information, when learning about themselves, learning about others, and making decisions.^{28,29,30} This may be due to negative information automatically activating threat responses—knowing about possible negative outcomes allows for planning and avoidance of potentially harmful or painful experiences^{31,32,33}.

Prior work has explored the role of negativity for driving online behavior. In particular, negative language in online content has been linked to user engagement, that is, sharing activities^{22,34,35,36,37,38,39}. As such, negativity embedded in online content explains the speed and virality of online diffusion dynamics (e.g., response time, branching of online cascades)^{7,34,35,37,39,40,41}. Further, online stories from social media perceived as negative garner more reactions (e.g., likes, Facebook reactions)^{42,43}. Negativity in news increases physiological activations⁴⁴, and negative news is more likely to be remembered by users^{45,46,47}. Some previous work has looked also into negativity effects for specific topics such as political communication and economics^{34,48,49,50,51,52}. Informed by this, we hypothesized an effect of negative words on online news consumption.

The majority of studies on online behavior are correlational^{34,35,36,38,39,40,41,42}, while laboratory studies take subjects out of their natural environment. As such, there is little work examining the causal impact of negative language on real-world news consumption. Here, we analyze data from the Upworthy Research Archive⁵³, a repository of news consumption data that are both applied and causal. Due to the structure of this dataset, we are able to test the causal impact of negative (and positive) language on news engagement in an ecologically rich online context. Moreover, our dataset is large-scale, allowing for a precise estimate of the effect size of negative words on news consumption.

Data on online news consumption was obtained from Upworthy, a highly influential media website founded in 2012 that used viral techniques to promote news articles across social media^{53,54}. Upworthy has been regarded as one of the fastest growing media companies worldwide⁵³ and, at its peak, reached more users than established publishers such as the New York Times⁵⁵. Content was optimized with respect to user responses through data-driven methods, specifically randomized controlled trials (RCTs)⁵⁶. The content optimization by Upworthy profoundly impacted the media landscape (e.g., algorithmic policies were introduced by Facebook in response)²³. In particular, the strategies employed by Upworthy have also informed other content creators and news agencies.

Upworthy conducted numerous randomized control trials of news headlines on its website to evaluate the efficacy of differently worded headlines in generating article views⁵³. In each experiment, Upworthy users were randomly shown different headline variations for a news story, and user responses were recorded and compared. Editors were commonly required to propose 25 different headlines from which the most promising headlines were selected for experimental testing⁵⁷.

In the current paper, we analyze the effect of negative words on news consumption. Specifically, we hypothesize that the presence of negative words in a headline will increase the click-through rate for that headline. The Design Table is in Table 1. Using a text mining framework, we extract negative words and estimate the effect on click-through rate using a multilevel regression (see Methods). We provide empirical evidence from large-scale, randomized controlled trials in the field ($N = 22,743$). Overall, our data contains over 105,000 different variations of news headlines from Upworthy, which have generated ~5.7 million clicks and more than 370 million impressions.

In addition to examining the effect of negative words as our primary analysis, we further conduct a secondary analysis examining the effect of high and low arousal negative words. Negative sentiment consists of many discrete negative emotions. Prior work has proposed that certain discrete categories of negative emotions may be especially attention-grabbing⁵⁸. For example, high-arousal negative emotions such as anger or fear have been found to efficiently attract attention and be quickly recognizable in facial expressions and body language^{31,59,60}. This may be because of the social and informational value that high-arousal emotions like anger and fear holds – both could alert others in one’s group to threats, and paying preferential attention and recognition to these emotions could help the group survive^{27,32}. This may also be why in the current age, people are more likely to share and engage with online content that is embedding anger, fear, or sadness^{21,41,61,62}. Therefore, we examine the effects of words related to anger and fear (as high-arousal negative emotions), as well as sadness (as a low-arousal negative emotion). We also examine the effects of words related to joy (positive emotion), which we predict will be associated with lower click-through rates.

2 Results

The following analyses are based on a reserved portion of the data (the “confirmatory sample”), which was only made available after acceptance of a Stage 1 Registered Report. All pilot analyses (reported in the Stage 1 manuscript) were conducted on a subset of the total data, and have no overlap from the analyses for Stage 2. When reporting estimates, we abbreviate standard errors via “SE” and 99% confidence intervals via “CI”.

2.1 Randomized controlled trials comparing news consumption

Our dataset contains a total of $N = 22,743$ RCTs. These consist of $\sim 105,000$ different variations of news headlines from Upworthy.com that generated ~ 5.7 million clicks across more than 370 million overall impressions. After applying the pre-registered filtering procedure (see Methods), we obtained 12,448 RCTs. Each RCT compares different variations of news headlines that all belong to the same news story. For example, the headline “WOW: Supreme Court Have Made Millions Of Us Very, Very Happy” and “We’ll Look Back At This In 10 Years Time And Be Embarrassed As Hell It Even Existed” are different headlines used for the same story about the repeal of Proposition 8 in California. An average of 4.31 headline variations (median = 4) were tested in each RCT. The headline variations are then compared with respect to the generated click-through rate, defined as the ratio of clicks per impressions (see Table 2 for examples). Overall, the 12,448 RCTs comprise 53,699 different headlines, which received over 205 million impressions and 2,778,124 clicks.

In the experiments, the recorded click-through rate ranges from 0.00% to 14.89%. The average click-through rate across all experiments is 1.39%, and the median click rate was 1.07%. Furthermore, the distribution among click-through rates is right-skewed, indicating that only a

small proportion of news stories were associated with a high click-through rate (Figure 1A). For instance, 99% of headline variations have a click-through rate below 6%. The results lay the groundwork for identifying the drivers of high levels of news consumption (additional descriptive statistics are in Supplementary Table 1 and Supplementary Figure 1).

There are considerable differences between positive and negative language in news headlines (see Figure 1B). We find that positive words are more prevalent than negative words (Kolmogorov-Smirnov (KS) test: $D = 0.574$, $P < 0.001$, two-tailed). Overall, 2.83% of all words in news headlines are categorized as positive words, whereas 2.62% of all words are categorized as negative words. In our sample, the most common positive words are “love” ($n = 980$), “pretty” ($n = 746$), and “beautiful” ($n = 645$), and the most common negative words are “wrong” ($n = 728$), “bad” ($n = 588$), and “awful” ($n = 363$). Ninety percent (91.61%) of the news stories in our sample contain a headline with at least one positive or negative word (i.e., 11,403 out of a possible 12,448), and 63.55% of headlines contain at least one word from our dictionaries (i.e., 34,124 out of a possible 53,699). Further statistics with word frequencies are in Supplementary Table 2.

2.2 Effect of negative language on news consumption

Randomized controlled experiments were used to estimate the effect of positive and negative words on news consumption, that is, the click-through rate. We employed a multilevel binomial regression that accommodates a random effects specification to capture heterogeneity among news stories (for details, see Methods; coefficient estimates are in Supplementary Table 3).

Positive and negative language in news headlines are both important determinants of click-through rates (see Figure 2A–C). Consistent with the *negativity bias hypothesis*, the effect

for negative words is positive ($\beta = 0.015$, $SE = 0.001$, $z = 17.423$, $P < 0.001$, 99% CI = [0.013, 0.018]), suggesting that a larger proportion of negative words in the headline increases the propensity of users to access a news story. A one standard deviation larger proportion of negative words increases the odds of a user clicking the headline by 1.5%. For a headline of average length (14.965 words), this implies that, for each negative word, the click-through rate increases by 2.3%. In contrast, the coefficient for positive words is negative ($\beta = -0.008$, $SE = 0.001$, $z = -9.238$, $P < 0.001$, 99% CI = [-0.010, -0.006]), implying that a larger proportion of positive words result in fewer clicks. For each standard deviation increase in the proportion of positive words per headline, the likelihood of a click decreases by 0.8%. Put differently, for each positive word in a headline of average length, the click-through rate decreases by 1.0%.

The estimated effects hold when adjusting for length and text complexity. A longer news headline increases the click-through rate ($\beta = 0.040$, $SE = 0.001$, $z = 43.945$, $P < 0.001$, 99% CI = [0.038, 0.043]). The click-through rate is decreased by a higher complexity score ($\beta = 0.004$, $SE = 0.001$, $z = -4.163$, $P < 0.001$, 99% CI = [-0.006, -0.001]), albeit to a smaller extent. This finding implies that lengthier and less complex formulations are appealing to users and lead to higher levels of news consumption. The control for platform age is negative ($\beta = -0.309$, $SE = 0.005$, $z = -56.917$, $P < 0.001$, 99% CI = [-0.323, -0.295]). Hence, stories published later in Upworthy's career had lower click-through rates than stories published at the beginning of Upworthy's career, implying that Upworthy headlines were most successful when its editorial practices were novel to online users.

2.3 Regression analysis with varying-slopes

Following our pre-registration, we further report results from a regression analysis with random effects and additional varying-slopes in the sentiment variables (Table 3). As such, the receptivity to language is no longer assumed to be equal across all experiments but is allowed to vary. Again, the coefficients are negative for positive words and positive for negative words. This thus implies that positive language decreases the clickability of news headlines, while negative language increases it. Furthermore, this is consistent with the analysis based on a random effects model without varying-slopes. Importantly, the results from both the varying-slopes model and our main model align with our pre-registered hypotheses, providing converging evidence of a negativity bias in news consumption.

Altogether, we find that a higher share of negative language in news headlines increases the click-through rate, whereas a higher share of positive language decreases the click-through rate. It is important to note that headlines belong to the *same* news story, and, therefore, phrasing news—regardless of its story—in a negative language increases the rate of clicking on a headline.

2.4 Robustness checks

The robustness of our preliminary analysis was confirmed by a series of further checks (see Supplement D). First, we repeated the analysis with alternative sentiment dictionaries as an additional validation. Specifically, we examined both the NRC dictionary^{64,65} and the SentiStrength dictionary⁶⁶. We found that the coefficient estimates were in good agreement, contributing to the robustness of our results (see Supplementary Table 4 & Supplementary Figure 2). Second, we repeated the above main analysis with an alternative approach for negation handling (i.e., a different neighborhood for inverting the polarity of words). This approach led to

qualitatively identical results (see Supplementary Table 5 and Supplementary Figure 3). Third, we repeated the analyses above using alternate text complexity measures. We found that there is a robust effect for both positive and negative words (see Supplementary Table 6 and Supplementary Figure 4). Fourth, we controlled for quadratic effects. We still observed a dominant effect of negative language (see Supplementary Table 7 and Supplementary Figure 5). Fifth, we repeated the analyses above but removed headlines where both positive and negative words were simultaneously present. As such, we end up with all headlines that exclusively include either positive or negative words. We found that headlines with negative words continued to be more likely to be clicked on than headlines with positive words (see Supplementary Table 9). Sixth, we repeated the same analyses as above, but removed all image RCTs where the teaser images were varied. This approach led to nearly identical results (see Supplementary Table 9). Seventh, we computed a single sentiment score, which is given by the net difference between the proportion of positive words and the proportion of negative words. As expected, negative sentiment increased click-through rate (see Supplementary Table 10). All aforementioned robustness checks were included in the Stage 1 Analysis Plan. During Stage 2 revisions, one reviewer asked for us to check if the results were robust when the CTR was log-transformed because of the positive skew of the data. While this analysis was not registered in our Stage 1 manuscript, we found that negative sentiment continued to increase click-through rates (see Supplementary Table 11).

We investigated moralized language as a possible moderator of positive and negative language in driving the click-through rate. Previously, moralized language was identified as an important driver of the diffusion of social media content³⁹. We extended the regression models from our main analysis with interaction terms between the proportion of moral words per

headline and the variables for the proportion of positive and negative words. In addition, we included the proportion of moral words per headline as a regressor to estimate its direct effect. We found a negative and statistically significant direct effect of moralized language on click-through rate ($\beta = -0.024$, $SE = 0.001$, $z = -17.067$, $P < 0.001$, 99% CI = [-0.028, -0.020]) and negative and statistically significant effects for the interactions between the proportion of moral words and the proportion of positive ($\beta = -0.006$, $SE = 0.001$, $z = -5.321$, $P < 0.001$, 99% CI = [-0.010, -0.003]) and negative words ($\beta = -0.007$, $SE = 0.001$, $z = -7.048$, $P < 0.001$, 99% CI = [-0.010, -0.005]). Full model results are in Supplementary Table 12 and Supplementary Figure 6. The direct negative effect of moral words suggests that headlines that contained moral words were less likely to be clicked on. The interactions suggest that positive words have a more negative effect and negative words have a less positive effect when moral words are also present in the headline. The results thus point towards a direct and a moderating role of moralized language. Yet even when controlling for moralized language, the direct effect of negative language was still present and continues to support the *negativity bias hypothesis*.

2.5 Negativity effect across different news topics

We examined the effect of negative language across various news topics. The rationale is that news stories in our data comprise various topics, for which the effect of negativity on the click-through rate could potentially differ. To this end, we applied topic modeling as in earlier research (e.g., Refs. ^{22,67}). Topic modeling infers a categorization of large-scale text data through a bottom-up procedure, thereby grouping similar content into topics. In order to obtain topic labels for each headline, we employed the topic model from our pre-registration phase (see Stage 1 of the Registered Report) that groups headlines into 7 topics: “Entertainment,” “Government & Economy,” “LGBT,” “Parenting & School,” “People,” and “Women Rights & Feminism.” We

applied this topic model to infer topic labels for each headline in our dataset. Characteristic words for each topic are shown in Supplementary Table 13, and representative headlines for each topic are shown in Supplementary Table 14.

The topic labels were validated whether they provide meaningful representations. In a user study, participants were shown headlines from each topic and were asked to select which topic the headline best fit into (topic intrusion test). Participants ($k = 10$) identified the topic from the correct headline in 51.1% of the cases. For comparison, a random guess would lead to an accuracy of 25%, implying that participants are roughly twice as good. This improvement over the random guess was further statistically significant ($\chi^2 = 249.61$, $P < 0.001$). Details are in Supplementary Table 15.

We found significant differences in the baseline click-through rate among topics. For this, we estimated a model where we additionally control for different topics via dummy variables, thus capturing the heterogeneity in how different topics generate clicks. Keeping everything else equal, we find that news generated more clicks when covering stories related to “Entertainment,” “LGBT,” and “People.” In contrast, news related to “Government & Economy” have a lower clickability (see Supplementary Table 16).

We then controlled for how the effect of negative language might vary across different topics. Here, we found that the variables of interest (i.e., the proportion of positive and negative words) significantly interact with different topics. For example, headlines relating to “Government and Economy,” “LGBT,” “Parenting and School,” and “People” received more clicks when they contained a large share of negative words. We also found that headlines relating to “LGBT,” “Life,” “Parenting and School,” and “People” received fewer clicks when they contained a large share of positive words. Overall, we found that negative language still has a

statistically significant positive effect on the click-through rate (see Supplementary Table 17). In sum, these results are consistent with the main analysis.

2.6 Extension to discrete emotions

We conducted secondary analyses examining the effects of discrete emotional words on the click-through rate. Summary statistics are in Supplementary Figure 7 and Supplementary Table 18. Prior work has suggested that certain discrete emotions such as anger^{38,41} may be particularly prevalent in online news. Furthermore, discrete emotions were found to be important determinants for various forms of user interactions (e.g., sharing^{36,37,38,39,40,41}), thus motivating that discrete emotions may also play a role for news consumption.

We report findings from four emotions (anger, fear, joy, sadness) for which we found statistically significant positive correlations between the human judgments of emotions and the dictionary scores (see Methods). We observed a statistically significant and positive coefficient for sadness ($\beta = 0.006$, $SE = 0.001$, $z = 5.295$, $P < 0.001$, 99% CI = [0.003, 0.009]), and a statistically significant negative effect for joy ($\beta = -0.009$, $SE = 0.001$, $z = -7.664$, $P < 0.001$, 99% CI = [-0.012, -0.006]) and fear ($\beta = -0.007$, $SE = 0.001$, $z = -5.919$, $P < 0.001$, 99% CI = [-0.009, -0.004]). A one standard deviation increase in sadness increases the odds of a user clicking the headline by 0.7%, while a one standard deviation increase in joy or fear decreases the odds of a user clicking on a headline by 0.9% and 0.7%, respectively. The coefficient estimate for anger ($\beta = 0.000$, $SE = 0.001$, $z = -0.431$, $P = 0.666$, 99% CI = [-0.003, 0.002]) was not statistically significant at common statistical significance thresholds.

We performed an equivalence test⁶⁸ to see if the relationship between the emotion score for anger and the CTR can be dismissed as null. The equivalence test involves defining a threshold

for the smallest meaningful effect (here: -0.001 to $+0.001$) and determining whether the effect of interest falls within that threshold. We find that the equivalence test for anger was undecided ($P = 0.399$, 99% CI = $[-0.003, 0.002]$), suggesting that the results are inconclusive with respect to a null effect. Our results for joy and sadness were consistent with our pre-registered hypotheses, while our results for anger and fear did not align with our hypotheses. Consistent with our previous findings, we observed that the click-through rate increases as the text length increases ($\beta = 0.037$, SE = 0.001, $z = 33.206$, $P < 0.001$, 99% CI = $[0.034, 0.040]$), and decreases as the complexity score increases ($\beta = -0.004$, SE = 0.001, $z = -3.410$, $P < 0.001$, 99% CI = $[-0.007, -0.001]$). Again, the click-through rate was lower for headlines at the end of Upworthy's career ($\beta = -0.312$, SE = 0.006, $z = -55.029$, $P < 0.001$, 99% CI = $[-0.327, -0.298]$). Full results are in Supplementary Table 19.

The above findings are supported by additional checks. First, we controlled for different topics in our regression model and, for this, utilized the previous categorization via topic modeling. When including topic dummies, we still found statistically significant positive effects for sadness and significant negative effects for fear and joy; the coefficient for anger was not statistically significant (see Supplementary Table 20). These results are thus consistent with the main analysis. For thoroughness, we also conducted exploratory analyses using the four other basic emotions from the NRC emotion dictionary for which the correlation with human judgments was considerably lower in our validation study (See Supplementary Tables 21 and 22). Here we observed a statistically significant negative effect on the click-through rate for anticipation. Furthermore, we studied the effects of 24 emotional dyads from Plutchik's model⁶⁹, which are complex emotions composed of two basic emotions⁷⁰. Consistent with findings from prior literature^{21,43,61}, we found that several dyads such as outrage and disapproval are associated

with higher click-through rates. The full results of our exploratory analyses on basic and higher-order emotions can be found in Supplementary Figures 8, 9, 10, & 11, and Supplementary Tables 23, 24, & 25.

3 Discussion

To examine the causal impact of emotional language on news consumption, we analyzed a large dataset⁵³ of more than 105,000 headlines that encompassed more than 370 million impressions of news stories from Upworthy.com. Consistent with our pre-registration, we find supporting evidence of a negativity bias hypothesis: news headlines containing negative language are significantly more likely to be clicked on, even after adjusting for the corresponding content of the news story. For a headline of average length (~15 words), the presence of a single negative word increased the CTR by 2.3%. In contrast, we find that news headlines containing positive language are significantly less likely to be clicked on. For a headline of average length, the presence of positive words in a news headline significantly decreases the likelihood of a headline being clicked on, by around 1.0%. The effects of positive and negative words are robust across different sentiment dictionaries, and after adjusting for length, text complexity, and platform age. While the observed effect sizes were noticeably smaller than in studies analyzing sharing behavior¹¹, increases/decreases of around 1-2% have been found to be meaningful differences when studying negativity bias in news consumption⁴⁵. Notably, we uncovered a negativity bias in the data even though Upworthy branded itself as a “positive news outlet”⁵³. Hence, while Upworthy readers may have chosen this outlet for its positive spin on the news, negative language increased news consumption, while positive language decreased it.

We compared the effect of negative words across different news topics. Our analyses revealed that the positive effect of negative words on consumption rates was strongest for news stories pertaining to “Government & Economy.” These results suggest that, consistent with prior work, individuals are especially likely to consume political and economic news when it is negative^{46,47,48,49,50,52}. Hence, people may be (perhaps unintentionally) selectively exposing

themselves to divisive political news, which ultimately may contribute to political polarization and intergroup conflict^{45,61,71} .

We further extended our analyses to examine the effect of discrete emotional language on news consumption. Consistent with our pre-registered hypotheses, basic emotions such as *sadness* increased the likelihood that a news headline was clicked on, while *joy* decreased the likelihood that a news headline was clicked on. Interestingly, we did not see a statistically significant effect for words related to *anger*, which prior research has suggested play a role in online diffusion^{22,38}. Furthermore, the effect of *fear* was significantly related to article consumption, but in the opposite direction of our pre-registered hypotheses. Interestingly, much work thus far has found that high valence emotions (like *fear* and *anger*) are related to online sharing behavior^{34,35,37,44}. Here, however, we measured online news consumption, a private behavior, which may account for differences from prior findings. Consumption behavior must be measured unobtrusively, so that it captures all content that individuals want to attend to, instead of what they want others to know they attended to. Sharing behavior is public, curated, and influenced by a myriad of social factors (e.g., signaling group identity, maintaining reputation)^{31,61,72}. The distinction between engaging with content online and consuming content online is important, as social considerations play a role in the decision to share content, but not necessarily to consume content⁶¹. Generally, people share only a fraction of the content they consume online, implying that engagement may be driven by different emotions and goals. These different motivations may make *fear* and *anger* more influential in the decision to share, compared to the decision to consume. Here, the Upworthy Research Archive gives us a large-scale opportunity to examine people's personal preference for news they attend to as

opposed to the news they want to share with others. Future research should investigate the differences between sharing behavior and consumption behavior (e.g., considering personality⁷³).

Upworthy's use of large-scale A/B testing allows social scientists the opportunity to analyze real-world behavior from millions of users in an experimental setting and thereby make causal inferences. Furthermore, Upworthy was extremely popular and continues to have a lasting impact on editorial practices through the media landscape^{23,56,57}. We found that headlines received fewer clicks when they were published later in Upworthy's career. This might suggest that Upworthy's "click-bait" editorial practices were most effective when they were novel to readers, and that the effectiveness decayed over time. This should make our data representative of news consumption across many contemporary online media sites. The results of our study demonstrate a robust and causal negativity bias in news consumption from a massive dataset from the field.

As with other research, ours is not free of limitations. Upworthy differs from more traditional news sources due to use of "click-bait" headlines. Nevertheless, we think that analyzing user behavior at Upworthy is important due to its large readership and influence on the media landscape. Moreover, it is important to note that we can only draw conclusions at the level of news stories but not at the level of individuals. Along similar lines, we are limited in the extent to which we can infer the internal state of a perceiver based on the language they write, consume, or share⁷⁴. In general, using online data to infer individual differences in subjective feelings is challenging. However, our analysis does not attempt to infer subject feelings; rather, we quantify how the presence of certain words is linked to concrete behavior. Hence, readers' preference for headlines containing negative words does not imply that users felt more negatively while reading said headlines. Our results show that negative words increase consumption rates, but make no claims regarding the subjective experience of readers.

Another potential limitation of these analyses is our use of discrete emotions via the NRC emotion lexicon. Our choice of the NRC emotion lexicon was due to two reasons: first, the NRC emotion lexicon is a prominent and comprehensive choice for examining discrete emotions in text⁶⁵, and, second, it captures various emotions at a granular level. In this regard, prior work has posited that specific emotions, such as *anger*, *outrage*, or *disapproval*, are prevalent in online news^{21,43,58,75}. Motivated by this, we felt it necessary to include an analysis of discrete emotions for comparability and richness. Nevertheless, following a psychological constructionist perspective, models of emotion involving a 2x2 dimensional space for valence and arousal have been proposed^{72,76}, yet the availability of corresponding dictionaries for such emotion models are limited, because of which we opted for the NRC emotion lexicon.

Understanding the biases that influence people's consumption of online content is critical, especially as misinformation, fake news, and conspiracy theories proliferate online. Even publishers marketed as "good news websites" are benefiting from negativity, demonstrating the need for a nuanced understanding of news consumption. Knowing what features of news makes articles interesting to people is a necessary first step for this purpose and will enable us to increase online literacy and to develop transparent online news practices.

4 Methods

4.1 Ethics information

The research complies with all relevant ethical regulations. Ethics approval (2020-N-151) for the main analysis was obtained from the institutional review board at ETH Zurich. For the user validation, ethics approval (IRB-FY2021-5555) was obtained from the institutional review board at New York University. Participants in the validation study were recruited from the subject pool of the Department of Psychology at New York University in exchange for 0.5 hrs of research credit for varying psychology courses. Participants provided informed consent for the user validation studies. New York University did not require IRB approval for the main analysis, as it is not classified as human subjects research.

4.2 Large-scale field experiments

In this research, we build upon data from the Upworthy Research Archive⁵³. The data have been made available through an agreement between Cornell University and Upworthy. We have access to this dataset upon the condition of following the procedure of a Registered Report. In Stage 1, we had access only to a subset of the dataset (i.e., the “exploratory sample”), based on which we conducted the preliminary analysis for pre-registering hypotheses. In Stage 2 (this manuscript), we had access to a separate subset of the data (i.e., the “confirmatory sample”) based on which tested the pre-registered hypotheses. Here, our analysis was based on data from $N = 22,743$ experiments (RCTs) collected on Upworthy between January 24, 2013 and April 14, 2015.

Each RCT corresponds to one news story, in which different headlines for the same news story were compared. Formally, for each headline variation j in an RCT i ($i = 1, \dots, N$), the following statistics were recorded: (1) the number of impressions, that is, the number of users to

whom the headline variation was shown ($impressions_{ij}$), and (2) the number of clicks a headline variation generated ($clicks_{ij}$). The click-through rate is then computed via $CTR_{ij} = \frac{clicks_{ij}}{impressions_{ij}}$. The experiments were conducted separately (i.e., only a single experiment was conducted at the same time for the entire website) so each test can be analyzed as independent of all other tests.⁵³ Examples of news headlines in the experiments are in Table 2. The Upworthy Research Archive contains data aggregated at the headline level and, thus, does not provide individual level data for users.

The data were subject to the following filtering. First, all experiments solely consisting of a single headline variation were discarded. Single headline variations exist because Upworthy conducted RCTs on features of their articles other than headlines, predominantly teaser images. In many RCTs where teaser images were varied, headlines were not varied at all (image data has not been made available to researchers by the Upworthy Research Archive, so we are unable to incorporate image RCTs into our analyses but we validated our findings as part of the robustness checks). Second, some experiments contain multiple treatment arms with identical headlines, which were merged into one representative treatment by summing their clicks and impressions. These occurred when images *and* headlines were involved in RCTs for the same story. This is relatively rare in the dataset, but for robustness checks regarding image RCTs (see Supplementary Table 9).

The analysis in the current Registered Report Stage 2 is based on the confirmatory sample of the dataset⁵³, which was made available to us only after pre-registration was conditionally accepted. In the previous pre-registration stage, we presented the results of a preliminary analysis based on a smaller, exploratory sample (see Registered Report Stage 1). Both were processed

using identical methodology. The pilot sample for our preliminary analysis comprised 4,873 experiments, involving 22,666 different headlines before filtering and 11,109 headlines after filtering, which corresponds to, on average, 4.27 headlines per experiment. On average, there were approximately 16,670 participants in each RCT. Additional summary statistics are given in Supplementary Table 1.

4.3 Design

We present a Design Table summarizing our methods in Table 1.

4.4 Sampling plan

Given our opportunity to secure an extremely large sample where the N was predetermined, we chose to run a simulation before pre-registration to estimate what level of power we would achieve for observing an effect size represented by a regression coefficient of 0.01 (i.e., a 1% effect on the odds of clicks from a standard deviation increase in negative words). This effect size is slightly more conservative than estimates of effect sizes from pilot studies (see Stage 1 of the Registered Report), and is derived from theory⁷⁸. The size of the confirmatory Upworthy data archive is $N = 22,743$ RCTs, with between three and twelve headlines per RCT. This thus corresponds to a total sample of between 68,229 and 227,430 headlines. Because we were not aware of the exact size during pilot testing, we generated datasets through a bootstrapping procedure that sampled $N = 22,743$ RCTs with replacement from our pilot sample of tests. We simulated 1000 such datasets and for each dataset we generated “clicks” using the estimated parameters from the pilot data. Finally, each dataset was analyzed using the model as described. This procedure was repeated for both models (varying-intercepts, and a combination of varying-intercepts and varying-slopes). We found that, under the assumptions of effect size,

covariance matrix and data generating process from our pilot sample, we will have greater than 99% power to detect an effect size of 0.01 in the final sample for both models.

4.5 Analysis plan

4.5.1 Text mining framework

Text mining was used to extract emotional words from news headlines. To prepare the data for the text mining procedure, we applied standard preprocessing to the headlines. Specifically, the running text was converted into lower-case and tokenized, and special characters (i.e., punctuations and hashtags) were removed. We then applied a dictionary-based approach analogous to earlier research^{22,39,40,41}.

We performed sentiment analysis based on the Linguistic Inquiry and Word Count (LIWC)⁶³. The LIWC contains word lists classifying words according to both a positive ($n = 620$ words, e.g. “love” and “pretty”) and negative sentiment ($n = 744$ words, e.g. “wrong” and “bad”). A list of the most frequent positive and negative words in our dataset are in Supplementary Table 2.

Formally, sentiment analysis was based on single words (i.e., unigrams) due to the short length of the headlines (mean length: 14.965 words). We counted the number of positive words ($n_{positive}$) and the number of negative words ($n_{negative}$) in each headline. A word was considered “positive” if it is in the dictionary of positive words (and, vice versa, for “negative” words). We then normalized the frequency by the length of the headline, that is, the total number of words in the headline (n_{total}). This yielded the two separate scores

$$Positive_{ij} = \frac{n_{positive}}{n_{total}} \text{ and } Negative_{ij} = \frac{n_{negative}}{n_{total}},$$

for headline j in experiment i . As such, the corresponding scores for each headline represent percentages. For example, if a headline has 10 words out of which one is classified as “positive” and none as “negative,” the scores are $Positive_{ij} = 10\%$ and $Negative_{ij} = 0\%$. If a headline has 10 words and contains one “positive” and one “negative” word, the scores are $Positive_{ij} = 10\%$ and $Negative_{ij} = 10\%$. A headline may contain both positive and negative words, so both variables were later included in the model.

Negation words (e.g., “not,” “no”) have the ability to invert the meaning of statements and thus the corresponding sentiment. We performed negation handling as follows. First, the text was scanned for negation terms using a predefined list, and then all positive (or negative) words in the neighborhood were counted as belonging to the opposite word list, i.e., they were counted as negative (or positive) words. In our analysis, the neighborhood (i.e., the so-called negation scope) was set to 3 words after the negation. As a result, a phrase like “not happy” was coded as negative rather than positive. Here, we used the implementation from the *sentimentr* package (details: <https://cran.r-project.org/web/packages/sentimentr/readme/README.html>).

Using the above dictionary approach, our objective was to quantify the presence of positive and negative words. As such, we did not attempt to infer the internal state of a perceiver based on the language they write, consume, or share⁷⁴. Specifically, readers’ preference for headlines containing negative words does not imply that users *felt* more negatively while reading said headlines. In contrast, we quantified how the presence of certain words is linked to concrete behavior. Following this, our pre-registered hypotheses test whether negative words increase consumption rates (see Table 1).

We validated the dictionary approach in the context of our corpus based on a pilot study⁷⁹. Here, we used the positive and negative word lists from LIWC⁶³ and performed negation handling as described above. Perceived judgments of positivity and negativity in headlines correlate with the number of negative and/or positive words each headline contains. Specifically, we correlated the mean of the 8 human judges' scores for a headline with NRC sentiment rating for that headline. We found a moderate but significant positive correlation ($r_s = 0.303$, $P < 0.001$). These findings validate that our dictionary approach captures significant variation in the perception of emotions in headlines from perceivers. More details are available in Supplementary Tables 21 and 22.

Two additional text statistics were computed: First, we determined the length of the news headline as given by the number of words. Second, we calculated a text complexity score using the Gunning-Fog index⁸⁰. This index estimates the years of formal education necessary for a person to understand a text upon reading it for the first time: $0.4 \times (ASL + 100 \times n_{\text{wsy} \geq 3} / n_w)$, where ASL is the average sentence length (number of words), n_w is the total number of words, and $n_{\text{wsy} \geq 3}$ is the number of words with three syllables or more. A higher value thus indicates greater complexity. Both headline length and the complexity score were used as control variables in the statistical models. Results based on alternative text complexity scores are reported as part of the robustness checks.

The above text mining pipeline was implemented with the software tool R 4.0.2 using the packages *quanteda* (version 2.0.1) and *sentimentr* (version 2.7.1) for text mining.

4.5.2 Empirical model

We estimated the effect of emotions on online news consumption using a multilevel binomial regression. Specifically, we expected that negative language in a headline affects the probability that users click on a news story to access its content. To test our hypothesis, we specified a series of regression models where the dependent variable is given by the click-through rate.

We modeled news consumption as follows. Let $i = 1, \dots, N$ refer to the different experiments in which different headline variations for news stories are compared through an RCT. Let $clicks_{ij}$ denote the number of clicks from headline variation j belonging to news story i . Analogously, let $impressions_{ij}$ be the corresponding number of impressions. We followed prior approaches⁸⁰ and modeled the number of clicks to follow a binomial distribution via

$$clicks_{ij} \sim \text{Binomial}(impressions_{ij}, \theta_{ij}),$$

where $0 \leq \theta_{ij} \leq 1$ is the probability of a user clicking on a headline in a single Bernoulli trial and where θ_{ij} corresponds to the click-through rate of headline variation j from news story i .

We estimated the effect of positive and negative words on the click-through rate θ_{ij} and captured between-experiment heterogeneity through a multilevel structure. We further controlled for other characteristics across headline variations, namely length, text complexity, and the relative age of a headline (based on the age of the platform). The regression model is then given by

$$\begin{aligned} \text{logit}(\theta_{ij}) = & \alpha + \alpha_i + \beta_1 \text{Positive}_{ij} + \beta_2 \text{Negative}_{ij} \\ & + \gamma_1 \text{Length}_{ij} + \gamma_2 \text{Complexity}_{ij} + \gamma_3 \text{PlatformAge}_{ij}, \end{aligned}$$

where α is the global intercept and α_i is an experiment-specific intercept (i.e., random effect). Both α and α_i are assumed to be independent and identically normally distributed with a mean of zero. The latter captures heterogeneity at the experiment level; that is, some news stories might be more interesting than others. In addition, we controlled for the length ($Length_{ij}$) and complexity ($Complexity_{ij}$) of the text in the news headline, as well as the relative age of the current experiment with regard to the platform ($PlatformAge_{ij}$). The latter denotes the number of days of the current experiment since the first experiment on Upworthy.com in 2012 and thus allowed us to control for potential learning effects as well as changes in editorial practices over time. The coefficient β_2 is our main variable of interest: it quantifies the effect of negative words on the click-through rate.

In the above analysis, all variables were z -standardized for better comparability. That is, prior to estimation, we subtracted the sample mean and divided by the standard deviation. Because of this, the regression coefficients β_1 and β_2 quantify changes in the dependent variable in standard deviations. This allowed us to compare the relative effect sizes across positive and negative words (as well as emotional words later). Due to the logit link, the odds ratio is $100 \times (e^\beta - 1)$, which gives the percentage change in the odds of success as a result of a standard deviation change in the independent variable. In our case, a successful event is indicated by the user clicking the headline.

The above regression builds upon a global coefficient for capturing the effect of language on click-through rate, and, as such, the language reception is assumed to be equal across different RCTs. This is consistent with previous works where a similar global coefficient (without varying-slopes) was used^{22,34,38,39}. However, there is reason to assume that the receptivity to language might vary across RCTs and thus among news (e.g., the receptivity of negative

language might be more dominant for political news than for entertainment news, or for certain news topics over others). As such, the variance in the estimated regression coefficients is no longer assumed to be exactly zero across experiments but may vary. To do so, we augmented the above random effects model by an additional varying-slopes specification. Here, a multilevel structure was used that accounts for the different levels due to the experiments $i = 1, \dots, N$. Specifically, the coefficients β_1 and β_2 capturing the effect of positive and negative words on click-through rate, respectively, are allowed to vary across experiments. Of note, a similar varying-slopes formalization was only used for the main analysis based on positive and negative language, and not for the subsequent extension to emotional words where it is not practical due to the fact that there would be comparatively few treatment arms in comparison to the number of varying-slopes.

Here, we conducted the analysis based on both models, that is, (i) the random effect model and (ii) the random effect model with additional varying-slopes. If the estimates from both models are in the same direction, this should underscore the overall robustness of the findings. If it is the case that estimated coefficients from the random effect model and the random effect, varying-slopes model contradict each other, both results are reported but precedence in interpretation was given to the latter due to its more flexible specification.

All models were estimated using the *lme4* package (version 1.1.23) in R.

4.5.3 Extension to discrete emotional words

To provide further insights into how emotional language relates to news consumption, we extended our text mining framework and performed additional secondary analyses. We were

specifically interested in the effect of different emotional words (anger, fear, joy, and sadness) on the click-through rate.

Here, our analyses were based on the NRC emotion lexicon due to its widespread use in academia and the scarcity of other comparable dictionaries with emotional words for content analysis^{64,65}. The NRC lexicon comprises 181,820 English words that are classified according to the 8 basic emotions of Plutchik’s emotion model⁶⁸. Basic emotions are regarded as universally recognized across cultures, and on this basis, more complex emotions can be derived^{70,82}. The 8 basic emotions computed via the NRC were anger, anticipation, joy, trust, fear, surprise, sadness, and disgust.

We calculated scores for basic emotions embedded in news headlines based on the NRC emotion lexicon⁶⁴. We counted the frequency of words in the text that belong to a specific basic emotion in the NRC lexicon (i.e., an 8-dimensional vector). A list of the most frequent emotional words in our dataset is given in Supplementary Table 18. Afterward, we divided the word counts by the total number of dictionary words in the text, so that the vector is normalized to sum to one across the basic emotions. Following this definition, the embedded emotions in a text might be composed of, for instance, 40% *anger* while the remaining 60% are *fear*. We omitted headline variations that do not contain any emotional words from the NRC emotion lexicon (since, otherwise, the denominator was not defined). Due to this extra filtering step, we obtained a final sample of 39,897 headlines. We again accounted for negations using the above approach in that the corresponding emotional words are not attributed to the emotion but skipped during the computation (as there is no defined “opposite” emotion).

As a next step, we validated the NRC emotion lexicon for the context of our study through a user study . Specifically, we correlated the mean of the 8 human judges’ scores for a

headline with NRC emotion rating for that headline. We found that, overall, both mean user judgments on emotions and those from the NRC emotion lexicon are correlated (r_s : 0.114, $P < 0.001$). Furthermore, mean user judgements for four basic emotions were significantly correlated, namely anger (r_s : 0.22, $P = 0.005$), fear (r_s : 0.29, $P < 0.001$), joy (r_s : 0.24, $P = 0.002$), and sadness (r_s : 0.30, $P < 0.001$, respectively). The four other basic emotions from the NRC emotion lexicon showed considerably lower correlation coefficients in the validation study, namely anticipation (r_s : -0.07, $P = 0.341$), disgust (r_s : 0.01, $P = 0.926$), surprise (r_s : -0.06, $P = 0.414$), and trust (r_s : 0.12, $P = 0.122$). Because of that, we did not pre-register hypotheses for them.

The multilevel regression was specified analogous to the model above but with different explanatory variables, i.e.,

$$\begin{aligned} \text{logit}(\theta_{ij}) = & \alpha + \alpha_i + \beta_1 \text{Anger}_{ij} + \beta_2 \text{Fear}_{ij} + \beta_3 \text{Joy}_{ij} \\ & + \beta_4 \text{Sadness}_{ij} + \gamma_1 \text{Length}_{ij} + \gamma_2 \text{Complexity}_{ij} + \gamma_3 \text{PlatformAge}_{ij}, \end{aligned}$$

where α and α_i represent the global intercept and the random effects, respectively. Specifically, α is again the global intercept and α_i captures the heterogeneity across experiments $i = 1, \dots, N$. As above, we included the control variables, i.e., length, text complexity, and platform age. The coefficients β_1, \dots, β_4 quantify the effect of the emotional words (i.e., anger, fear, joy, and sadness) on the click-through rate.

Again, all variables were z -standardized for better comparability (i.e., we subtract the sample mean and divide by the standard deviation). As a result, the regression coefficients quantify changes in the dependent variable in standard deviations. This allows us to compare the relative effect sizes across different emotions.

Additional information

Data availability. The full data from the randomized controlled experiments in the field is available through the Upworthy Research Archive⁵³ (DOI: <https://doi.org/10.1038/s41597-021-00934-7>). The data used in the present manuscript is available at the following link: <https://osf.io/uscpf/>. The LIWC dictionary⁶³ is available for purchase (<https://www.liwc.app/>) . The NRC emotion lexicon⁶⁴ is publicly available for download (<https://nrc.canada.ca/en/research-development/products-services/technical-advisory-services/sentiment-emotion-lexicons>). In our analysis, we used the built-in version from the “sentimentr” package.

Code availability. Code that supports the findings of our study is available at the following link: <https://osf.io/uscpf/>

Protocol registration. The Stage 1 protocol for this Registered Report was accepted in principle on 11 March 2022. The protocol, as accepted by the journal, can be found at https://springernature.figshare.com/articles/journal_contribution/Negativity_drives_online_news_consumption_Registered_Report_Stage_1_Protocol_/19657452.

Author information

Acknowledgements. We are grateful to a John Templeton Foundation Grant (#61378) and a Russell Sage Foundation Grant (G-2110-33990) that funded J.V.B. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We thank Upworthy, as well as J. Nathan Matias, Kevin Munger, Marianne Aubin Le Quere, and Charles Ebersole for making the data available. We thank W. Brady for his helpful feedback.

Author contributions. C.R., N.P., K.S., P.P., J.V.B., and S.F. conceived and designed the experiments. C.R., N.P., and K.S. analyzed the data. C.R., N.P., K.S., P.P., J.V.B., and S.F. wrote the paper.

Competing interests. The authors declare no financial or non-financial competing interests.

Correspondence. Jay J. Van Bavel (jay.vanbavel@nyu.edu) and Stefan Feuerriegel (feuerriegel@lmu.de).

Tables

Question	Hypothesis	Sampling plan	Analysis plan	Interpretation given to different outcomes
<p>How does the presence of negative and positive words affect the click-through rate for news headlines?</p>	<p>The presence of negative words in a headline will significantly increase the click-through rate for that headline. The presence of positive words in a headline will significantly decrease the click-through rate for that headline.</p>	<p>A power analysis suggested that the sample size of the confirmatory dataset (22, 743 RCTs) will have sufficient power to achieve 99% power to detect an effect size of 0.01, considered to be a small effect size⁷⁷. This effect size is slightly more conservative than estimates of effect sizes from pilot studies, and is derived from theory⁷⁸.</p>	<p>We conducted a multilevel binomial model examining the effects of the proportion of negative words in a headline on the click-through rate, adjusting for the proportion of negative words in a headline, the number of positive words in the headline, the complexity of the headline as measured by the Gunning Fog Index, and the age of the story relative to the age of the Upworthy platform. We included random effects grouped by RCT, and used two models to test our hypothesis. One allows the intercept to vary, the other also allows the slope of negative words to vary.</p>	<p>A significant positive coefficient for negative words is interpreted as evidence that a higher ratio of negative words in a headline is associated with a greater click-through rate. A significant negative coefficient for negative words is interpreted as evidence that a higher ratio of negative words in a headline is associated with a lower click-through rate. A significant positive coefficient for positive words is interpreted as evidence that a higher ratio of positive words in a headline is associated with a greater click-through rate. A significant negative coefficient for positive words is interpreted as evidence that a higher ratio of positive words in a headline is associated with a lower click-through rate. We consider evidence to be conclusive only in cases where both models fit to the data agree in their qualitative conclusions about the effect of negative words.</p> <p>To evaluate effects where we cannot reject the null hypothesis, we will test for equivalence⁶⁸ against an interval of [-0.001, 0.001]. If our observed confidence interval is fully contained in this interval we will consider this evidence for a null effect, otherwise we</p>

				will consider the results inconclusive with respect to the null.
How does the presence of discrete emotional words affect the click-through rate for news headlines?	The presence of anger, fear, and sadness words in a headline will significantly increase the click-through rate for that headline. The presence of joy words in a headline will significantly decrease the click-through rate for that headline.	A power analysis suggested that the sample size of the confirmatory dataset (22,743 RCTs) will have sufficient power to achieve 99% power to detect effect sizes of 0.01.	We conducted a multilevel binomial model examining the effects of the four emotions (anger, fear, joy, and sadness) on the click-through rate, adjusting for the number of words in the headline, the complexity of the headline as measured by the Gunning Fog Index, and the age of the story relative to the age of the Upworthy platform. We included the RCT as a random intercept.	<p>A positive value for each of the emotions signifies a larger proportion of emotional words from that emotion in a headline. Therefore, a significant positive coefficient for the emotion is interpreted as evidence that headlines containing a word from the emotion (i.e., anger, fear, joy, and sadness) is associated with a greater click-through rate. Conversely, a negative value for each of the coefficients signifies that the proportion of emotional words from the emotion is more prevalent in a headline. Therefore, a significant negative coefficient for the emotion indicates that headlines containing a word from emotion (i.e., anger, fear, joy, and sadness) is associated with a smaller click-through rate.</p> <p>To evaluate effects where we cannot reject the null hypothesis, we tested for equivalence⁶⁸ against an interval of [-0.001, 0.001]. If our observed confidence interval is fully contained in this interval we will consider this evidence for a null effect, otherwise we consider the results inconclusive with respect to the null.</p>

Table 1: Design table.

#	Headline Variation	CTR
1	If The Numbers 4 And 20 Mean Something To You, You're Gonna Want To Hear This Shit	0.94%
2	What He Has To Say About Pot Is Going To Make Both Sides Angry , But Here He Goes	0.79%
3	Lots Of Things In Live Have Both A Benefit And A Harm , So Why Do We Only Obsess About This One?	0.60%
4	He Explains Why The Question 'What Are You Smoking' Is Actually Kind of Important .	0.58%
1	IMAGINE: You're Raped At Your Job And Your Boss Intentionally Tries To Shut You Up	0.92%
2	12 Minutes. If You Support Our Troops, Sacrifice at Least That Much To Them.	0.21%
1	Spoofers Set Up A Fake Agency To Show How Ridiculous Some People Are When It Comes To Immigrants.	0.65%
2	Something's Been Missing From Our Favorite Superhero Stories, And It Makes Reality Seem Kinda Silly	0.56%
3	Some Comic Book Lovers Might Need To Check Their Politics When They See What These Guys Have In Mind	0.53%
4	A New Agency Wants To Get Rid Of All Our Favorite Superheroes. I Laughed When I Saw Why.	0.41%
1	I Knew Which One She's Pick, But It Still Crushed Me	1.10%
2	First She Points To The Pretty Child. Then To The Ugly Child. Then My Heart Breaks.	0.85%
3	1 Little Girl, 5 Cartoons, And 1 Heartbreaking Answer	0.83%
4	What She Says About These Cartoons Says Something Incredibly Troubling About The World We Live In	0.66%

Table 2: Example experiments (randomized control trials) performed by Upworthy. Shown are four experiments and each with different headline variations subject to testing. Columns report the click-through rate (CTR) and the positive (in bold) and negative words (in italics) as classified by the LIWC dictionary.

	Coef	Lower CI	Upper CI	<i>P</i> -value
<i>Positive</i>	-0.017	-0.023	-0.010	< 0.001
<i>Negative</i>	0.018	0.011	0.025	< 0.001
<i>Length</i>	0.044	0.041	0.047	< 0.001
<i>Complexity</i>	-0.002	-0.005	0.001	0.1
<i>PlatformAge</i>	-0.311	-0.325	-0.296	< 0.001
(Intercept)	-4.491	-4.506	-4.476	< 0.001

Observations: 53,699

Table 3: Results for the regression model with varying-slopes for the proportion of positive and negative words. Experiment-specific intercepts and slopes for the sentiment variables (i.e., random effects) are also included. Reported are standardized coefficient estimates and 99% confidence intervals (CIs). *P*-values are calculated using two-sided *z*-tests. $N = 53,699$ headlines were examined over 12,448 RCTs.

Figure Legends:

Figure 1: **(A)** Complementary cumulative distribution function (CCDF) comparing the click-through rate across all headline variations ($N = 53,669$). **(B)** CCDF comparing the distribution of the ratio of positive and negative words across all headline variations ($N = 53,669$). Positive words are more prevalent than negative words. A Kolmogorov-Smirnov (KS) test shows that this difference is statistically significant ($P < 0.001$, two-tailed). The y-axes of both plots are on a logarithmic scale.

Figure 2: The effect of positive and negative words in news headlines ($N = 53,669$) on the click-through rate headlines were examined over 12,448 RCTs. **(A)** Shown are the estimated standardized coefficients (circles) with 99% confidence intervals (error bars) for positive and negative words and for further controls. The variable *PlatformAge* is included in the model during estimation but not shown for better readability. Full estimation results are in Supplementary Table 3. **(B,C)** Predicted marginal effects on the click-through rate (lines). The error bands (shaded area) correspond to 99% confidence intervals. Boxplots show the distribution of the variables in our sample (center line gives the median; box limits are upper and lower quartiles; whiskers denote minimum/maximum; points are outliers defined as being beyond 1.5x of the interquartile range).

References

- [1] Pooley, E.. Grins, Gore and Videotape: The Trouble with Local TV News. *New York Magazine* **22**, 36-44 (1989).
- [2] Pew. Americans almost equally prefer to get local news online or on TV set. *Pew Research Center's Journalism Project* (2019)
- [3] Olmstead, K., Mitchell, A., & Rosenstiel, T. Navigating news online: Where people go, how they get there and what lures them away. *Pew Research Center's Project for Excellence in Journalism* (2011).
- [4] Simon, H. A. Designing Organizations for an Information-Rich World. In M. Greenberger (Ed.), *Computers, Communications, and the Public Interest*. Baltimore, MD: The Johns Hopkins Press. 38-72. (1971).
- [5] Flaxman, S., Goel, S. & Rao, J. M. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly* **80**, 298–320 (2016).
- [6] Schmidt, A. L. et al. Anatomy of news consumption on Facebook. *Proceedings of the National Academy of Sciences (PNAS)* **114**, 3035–3039 (2017).
- [7] Bakshy, E., Messing, S. & Adamic, L. A. Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**, 1130–1132 (2015).
- [8] Allen, J., Howland, B., Mobius, M., Rothschild, D. & Watts, D. J. Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances* **6**, eaay3539 (2020).

- [9] Yang, T., Majó-Vázquez, S., Nielsen, R. K. & González-Bailón, S. Exposure to news grows less fragmented with an increase in mobile access. *Proceedings of the National Academy of Sciences (PNAS)* **117**, 28678–28683 (2020).
- [10] Godes, D. & Mayzlin, D. Using online conversations to study word-of-mouth communication. *Marketing Science* **23**, 545–560 (2004).
- [11] Berger, J. & Schwartz, E. M. What drives immediate and ongoing word of mouth? *Journal of Marketing Research* **48**, 869–880 (2011).
- [12] Antweiler, W. & Frank, M. Z. Is all that talk just noise? The information content of Internet stock message boards. *The Journal of Finance* **59**, 1259–1294 (2004).
- [13] Forbes. Can 'fake news' impact the stock market? URL <https://www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market/>. (2017).
- [14] Bollen, J., Mao, H. & Zeng, X. Twitter mood predicts the stock market. *Journal of Computational Science* **2**, 1–8 (2011).
- [15] Garfin, D. R., Silver, R. C. & Holman, E. A. The novel coronavirus (COVID-2019) outbreak: Amplification of public health consequences by media exposure. *Health Psychology* **39**, 355–357 (2020).
- [16] Aral, S. & Eckles, D. Protecting elections from social media manipulation. *Science* **365**, 858–861 (2019).
- [17] Bond, R. M. et al. A 61-million-person experiment in social influence and political mobilization. *Nature* **489**, 295–298 (2012).

- [18] Jones, J. J., Bond, R. M., Bakshy, E., Eckles, D. & Fowler, J. H. Social influence and political mobilization: Further evidence from a randomized experiment in the 2012 U.S. presidential election. *PLOS ONE* **12**, e0173851 (2017).
- [19] Levy, R. Social media, news consumption, and polarization: Evidence from a field experiment. *American Economic Review* **111**, 831-70. (2020).
- [20] Klein, E. *Why We're Polarized*. (Avid Reader Press / Simon & Schuster, New York, NY, 2020)
- [21] Crockett, M. J. Moral outrage in the digital age. *Nature Human Behaviour* **1**, 769–771 (2017).
- [22] Vosoughi, S., Roy, D., & Aral, S. The spread of true and false news online. *Science* **359**, 1146–1151 (2018).
- [23] Sanders, S. Upworthy was one of the hottest sites ever. You won't believe what happened next. *NPR*
<https://www.npr.org/sections/alltechconsidered/2017/06/20/533529538/upworthy-was-one-of-the-hottest-sites-ever-you-wont-believe-what-happened-next> (2017).
- [24] Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. Bad is stronger than good. *Review of General Psychology* **5**, 323-370 (2001).
- [25] Rozin, P., & Royzman, E. B. Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review* **5**, 296-320 (2001).

- [26] Carver, L. J., & Vaccaro, B. G. 12-Month-old infants allocate increased neural resources to stimuli associated with negative adult emotion. *Developmental Psychology* **43**, 54–69 (2007).
- [27] Dijksterhuis, A., & Aarts, H. On wildebeests and humans: The preferential detection of negative stimuli. *Psychological Science* **14**, 14–18 (2003).
- [28] Müller-Pinzler, L., Czekalla, N., Mayer, A. V., Stolz, D. S., Gazzola, V., Keysers, C., Paulus, F. M., & Krach, S. Negativity-bias in forming beliefs about own abilities. *Scientific Reports* **9**, 14416 (2019).
- [29] Boydston, A. E., Ledgerwood, A., & Sparks, J. A negativity bias in reframing shapes political preferences even in partisan contexts. *Social Psychological and Personality Science*, **10**, 53–61 (2019).
- [30] Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology* **75**, 887–900 (1998).
- [31] Öhman, A., & Mineka, S. Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review* **108**, 483–522 (2001).
- [32] Öhman, A., Flykt, A., & Esteves, F. Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General* **130**, 466–478 (2001).
- [33] Shoemaker, P. J. Hardwired for news: Using biological and cultural evolution to explain the surveillance function. *Journal of Communication* **46**, 32–47 (1996).

- [34] Stieglitz, S. & Dang-Xuan, L. Emotions and information diffusion in social media: Sentiment of microblogs and sharing behavior. *Journal of Management Information Systems* **29**, 217–248 (2013).
- [35] Naveed, N., Gottron, T., Kunegis, J. & Alhadi, A. C. Bad news travel fast: A content-based analysis of interestingness on Twitter. In *Proceedings of the 3rd International Web Science Conference* (2011).
- [36] Kim, J. & Yoo, J. Role of sentiment in message propagation: Reply vs. retweet behavior in political communication. In *International Conference on Social Informatics*, 131–136 (2012).
- [37] Berger, J., & Milkman, K. L. What makes online content viral? *Journal of Marketing Research* **49**, 192-205 (2012).
- [38] Chuai, Y. & Zhao, J. Anger makes fake news viral online. <https://arxiv.org/pdf/2004.10399> (2020).
- [39] Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences (PNAS)* **114**, 7313–7318 (2017).
- [40] Pröllochs, N., Bär, D., & Feuerriegel, S. Emotions explain differences in the diffusion of true vs. false social media rumors. *Scientific Reports* **11**, 22721 (2021).
- [41] Pröllochs, N., Bär, D., & Feuerriegel, S. Emotions in online rumor diffusion. *EPJ Data Science*, **10**, 51 (2021).

- [42] Zollo, F., Novak, P. K., Del Vicario, M., Bessi, A., Mozetič, I., Scala, A., Caldarelli, G., & Quattrociocchi, W.. Emotional dynamics in the age of misinformation. *PLOS ONE* **10**, e0138740. (2015).
- [43] Rathje, S., Van Bavel, J. J., & van der Linden, S. Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences* **118** (2021).
- [44] Soroka, S., Fournier, P., & Nir, L. Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences (PNAS)* **116**, 18888–18892 (2019).
- [45] Trussler, M., & Soroka, S. Consumer demand for cynical and negative news frames. *The International Journal of Press/Politics* **19**, 360–379 (2014).
- [46] Meffert, M. F., Chung, S., Joiner, A. J., Waks, L., & Garst, J. The effects of negativity and motivated information processing during a political campaign. *Journal of Communication* **25**, 27–51 (2006).
- [47] Bradley, S. D., Angelini, J. R., & Lee, S. Psychophysiological and memory effects of negative political ADS: Aversive, arousing, and well remembered, *Journal of Advertising* **36**, 115–127 (2007).
- [48] Soroka, S., & McAdams, S. News, politics, and negativity. *Political Communication* **32**, 1–22 (2012)
- [49] Lengauer, G., Esser, F., & Berganza, R. Negativity in political news: A review of concepts, operationalizations and key findings. *Journalism* **13**, 179–202 (2012).

- [50] Jang, S. M., & Oh, Y. W. Getting attention online in election coverage: Audience selectivity in the 2012 US presidential election. *New Media & Society* **18**, 2271–2286 (2016).
- [51] Haselmayer, M., Meyer, T. M., & Wagner, M. Fighting for attention: Media coverage of negative campaign messages. *Party Politics* **25**, 412–423 (2019).
- [52] Soroka, S. N. Good News and Bad News: Asymmetric Responses to Economic Information. *The Journal of Politics* **68**, 372–385 (2006).
- [53] Matias, J., Munger, K., Le Quere, M.A., Ebersole, C. The Upworthy Research Archive, a time series of experiments in U.S. media. *Nature Scientific Data* **8**, 195 (2021)
- [54] Karpf, D. *Analytic Activism: Digital Listening and the New Political Strategy*. Oxford Studies in Digital Politics (Oxford University Press, New York, NY, 2016).
- [55] Thompson, D. I thought I knew how big Upworthy was on Facebook: Then I saw this. *The Atlantic*.
<https://www.theatlantic.com/business/archive/2013/12/i-thought-i-knew-how-big-upworthy-was-on-facebook-then-i-saw-this/282203/> (2012).
- [56] Fitts, A. S. The king of content: How Upworthy aims to alter the web, and could end up altering the world. *Columbia Journalism Review* **52** (2014).
- [57] Upworthy. How to make that one thing go viral, *SlideShare*
<https://www.slideshare.net/Upworthy/how-to-make-that-one-thing-go-viral-just-kidding/25> (2012).

- [58] Soroka, S., Young, L., & Balmas, M. Bad news or mad news? Sentiment scoring of negativity, fear, and anger in news content. *The ANNALS of the American Academy of Political and Social Science* **659**, 108–121 (2015).
- [59] Fox, Elaine, Victoria Lester, Riccardo Russo, R. J. Bowles, Alessio Pichler, and Kevin Dutton. Facial expressions of emotion: Are angry faces detected more efficiently? *Cognition & Emotion* **14**, 61–92 (2000)
- [60] De Gelder, B. Towards the neurobiology of emotional body language. *Nature Reviews Neuroscience* **7**, 242–249 (2006).
- [61] Brady, W. J., Crockett, M. J., & Van Bavel, J. J. The MAD model of moral contagion: The role of motivation, attention, and design in the spread of moralized content online. *Perspectives on Psychological Science* **15**, 978–1010 (2020).
- [62] Spring, V. L., Cameron, C. D., & Cikara, M.. The upside of outrage. *Trends in Cognitive Sciences* **22**, 1067–1069 (2018).
- [63] Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. The development and psychometric properties of LIWC2015. <http://hdl.handle.net/2152/31333> (2015).
- [64] Mohammad, S. & Turney, P. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 26–34 (2010).
- [65] Mohammad, Saif M. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. *Emotion Measurement*. 201–237, (2016).

- [66] Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology* **61**, 2544–2558 (2010).
- [67] Toetzke, M., Banholzer, N., & Feuerriegel, S.. Monitoring global development aid with machine learning. *Nature Sustainability* **5**, 533–541 (2022).
- [68] Lakens, D., Scheel, A. M., & Isager, P. M. Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science* **1**, 259–269 (2018).
- [69] Plutchik, R. *Emotion: Theory, research, and experience* (Academic Press, Orlando, FL, 1984), 2 edn.
- [70] Ekman, P. An argument for basic emotions. *Cognition & Emotion* **6**, 169–200 (1992).
- [71] Van Bavel, J. J., Rathje, S., Harris, E., Robertson, C., & Sternisko, A. How social media shapes polarization. *Trends in Cognitive Sciences* **25**, 913–916 (2021).
- [72] Barrett, L. F., & Russell, J. A. (eds.). *The psychological construction of emotion* (Guilford Publications, New York, NY, 2014).
- [73] Bachleda, S., Neuner, F. G., Soroka, S., Guggenheim, L., Fournier, P., & Naurin, E.. Individual-level differences in negativity biases in news selection. *Personality and Individual Differences* **155**, 109675 (2020).
- [74] Kross, E., Verduyn, P., Boyer, M., Drake, B., Gainsburg, I., Vickers, B., ... & Jonides, J. Does counting emotion words on online social networks provide a window into people's subjective experience of emotion? A case study on Facebook. *Emotion* **19**, 97–107 (2019).

- [75] Jakubik, J., Vössing, M., Bär, D., Pröllochs, N., & Feuerriegel, S. (2023). Online emotions during the storming of the US Capitol: Evidence from the social media network Parler. *International Conference on Web and Social Media (ICWSM)*.
- [76] Barrett, L. F. Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cognition & Emotion* **12**, 579–599 (1998).
- [77] Khalilzadeh, J., & Tasci, A. D. Large sample size, significance level, and the effect size: Solutions to perils of using big data for academic research. *Tourism Management* **62**, 89–96 (2017).
- [78] Cohen, J. *Statistical power analysis for the behavioral sciences*. (Lawrence Erlbaum Associates Publishers, Hillsdale, NJ, 2013), 2nd edn.
- [79] Song, H. et al. In validations we trust? The impact of imperfect human annotations as a gold standard on the quality of validation of automated content analysis. *Political Communication* **37**, 550–572 (2020).
- [80] Gunning, R. *The Technique of Clear Writing*. (McGraw-Hill, Toronto, CA, 1952).
- [81] Richardson, M., Dominowska, E. & Ragno, R. Predicting clicks: Estimating the click-through rate for new ads. In *Proceedings of the 16th International Conference on World Wide Web*, 521–530 (2007).
- [82] Sauter, D. A., Eisner, F., Ekman, P. & Scott, S. K. Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences (PNAS)* **107**, 2408–2412 (2010).