# How digital media markets amplify news sentiment*

Lara Marie Berger†

This version: January 15, 2024

*Preliminary draft.*

**Abstract**

Capturing attention through appealing headlines is far more important for news companies in digital than in analog media markets, but it is so far unclear if this changes news content. This paper provides evidence that this shift in incentives enhances the sentimental slant of news headlines. A comparison of online and offline versions of the same newspapers illustrates that headlines online are more often formulated emotionally. An experiment with professional journalists reveals that this difference can be at least partially explained by an increased incentive to generate attention: If journalists are compensated relative to the click rates their headlines receive, they significantly more often put headlines containing emotional words on top of a given article. A second experiment shows that such an amplification of sensationalist framing has economic implications in the short run, as emotional headlines can translate into emotional reactions and distortions in expectations of their readers.

**JEL Codes:** D83, D91, G41, L82

**Keywords:** Media Bias, News Sentiment, Digitization

*I thank Anna Kerkhof, Johannes Münster, Bettina Rockenbach, Christopher Roth and participants of the C-SEB early ideas workshop at University of Cologne, the ESA European Meeting 2022, the Workshop in Field Experiments in Economics and Business Heilbronn and the 2nd ECONtribute Young Economist Workshop Cologne for very helpful comments and suggestions. Nikola Noske provided excellent research assistance. I thank the Center for Social and Economic Behavior (C-SEB) of the University of Cologne for its financial support. The experiments were approved by the Ethics Committee of the University of Cologne (reference 210036LM). The experiments were pre-registered in the AEA Registry under registry number AEARCTR-0008658. Any mistake is my own.

†University of Cologne. E-mail: lara.berger@uni-koeln.de

# 1. Introduction

*"If it bleeds, it leads."* - This common phrase among journalists captures the tendency of many reporters to prioritize sensationalist and negative/violent content. Stories that evoke emotions are oftentimes *the lead*, which refers to the most prominent story on top of a page. In this way, media content is frequently making the world seem more sensational than it actually is.[1]

Interestingly, digitization shifts incentives in news markets in a way that might amplify this *sensationalist bias*, as attracting readers with news online works different than offline. For example, while in the offline world at most front-page headlines have to be catchy to increase sales, in the online sphere every headline is in competition with thousands of others. This increases incentives for formulating each headline as appealing as possible. Digitization also enables journalists to learn a lot about reading preferences of their online audiences. They can observe click-rates, reading duration or sharing rates on social media in real-time or even conduct small experiments to find out which headline performs best[2]. As there is sound evidence that humans react stronger to information that triggers emotional (and especially negative) arousal as opposed to neutral information[3], digitization might in this way induce journalists to write online even more emotional headlines than offline.

This paper provides descriptive and experimental evidence supporting this hypothesis. In a first step, I compare headlines in the online and offline versions of a wide range of newspapers. Fine-tuning a machine learning model to classify hundreds of thousands of headlines on economic issues of major German newspapers reveals that headlines in the online versions are 23 percent (0.45 standard deviations) more likely to be emotional than offline. This means headlines online are both more often positive and more often negative. However, negativity is dominant on average: When sentiment is regarded as a measure ranging from negative ($-1$) to positive ($+1$) tone, headlines online are on average 0.1 standard deviations more negative than offline. These findings are robust to the use of different classification algorithms and a hand-coded random subset of the data[4]. Furthermore, the difference in sentiment seems to be quite general. It is not driven by the political orientation of a newspaper, by whether it is a tabloid or a quality outlet or by targeting a regional or national audience. The results also reproduce in samples of headlines from non-economic news and from an US-American newspaper.

While this descriptive analysis is informative about the existence of a difference in tonality of headlines online and offline, it remains unclear whether it is actually *caused* by shifts in incentives for journalists. I conduct an experiment with professional journalists as participants (N=201) to shed light on this. The compensation of the journalists is randomly varied so that it resembles differences of incentives in *digital* media markets (treatment group) as opposed to *analogue* media

[1]This type of media slant has often been documented as negativity or positivity bias. Examples are Ryu (1982); Hofstetter and Dozier (1986); Goidel and Langley (1995); Farnsworth and Lichter (2011); Garz (2014); Soroka (2016); Soroka et al. (2018); Bleich and van der Veen (2021); Kayser and Peress (2021) and Soroka and Krupnikov (2021)). It is also inherent to human communication to prioritize stories that are in some way sensational or unusual, even though most things that happen are probably quite the opposite. For example, someone working in a bakery will most likely not tell their family in the evening who came by to buy bread. However, if they happen to have Will Smith as a customer, they will probably share this, even though Will Smith is not at all representative for their customers. The same logic applies to how the news industry frames and selects stories.

[2]See Leung and Strumpf (2022) for recent evidence on A/B-testing.

[3]See Rozin and Royzman (2001) for a review. Soroka et al. (2019) provide an application to reactions to news.

[4]This subset was was hand-coded by a research assistant

markets (control group). All journalists have to choose a headline with either positive, neutral or negative sentiment for a given article about forecasts of German GDP growth. Those in the treatment group are paid depending on how many readers choose to click on the headline of the article afterwards, whereas those in the control group receive a flat rate independently of the performance of their headline. I find that journalists are 20 percent more likely to choose an emotional (=either positive or negative) headline when they are paid per click. This corresponds to an effect size of 0.41 standard deviations.

To understand the relevance of this amplification of news sentiment in headlines it is important to know if it has any economic impact on readers. To learn more about this I consider the reactions of readers to the headlines chosen by the journalists in a second experiment. I randomly expose a sample of student participants (N=299) to each of the headlines and elicit clicking behavior, as well as emotional reactions, investment decisions and forward-looking economic expectations. I detect evidence for reactions in some of the considered measures. Readers of the positive headline state to feel significantly better than readers of the neutral or negative headline. Those who read an emotional headline have expectations about the economic development that are further away from the forecasts in the article than those who read the neutral headline, suggesting that they use the given information to update their expectations to a lower degree.

Overall, this paper demonstrates that the internet amplifies sensationalist media slant through a change in incentives for journalists. This slant has the potential to distort beliefs of readers, which might cause them to take sub-optimal decisions.

Thereby, this paper makes contributions to two streams of literature. The first is a body of work on negativity and positivity biases in the news media. Papers in this stream of literature typically document a distortion in the valence of news relative to some more objective measure[5] or illustrate the tendency of readers to react stronger to negative and positive news (as opposed to neutral or ambiguous content)[6].[7] However, only a few papers provide evidence on a how the tone of news headlines and economic incentives of a newspaper are related and, so far, all of the studies that do this are correlational. Arango-Kure et al. (2014) for example document an association of negative cover stories and magazine sales and Dertwinkel-Kalt et al. (2022) find a relationship between online click-rates and headline negativity. This paper complements these findings by providing the, to the best of my knowledge, first causal evidence for this relationship.

Additionally, evidence on the effects of this type of media slant on readers is so far non-existent. While the information provision literature[8] and the literature on the influence of sentiment on financial markets[9] suggest that such effects are very plausible, the paper at hand can directly show

[5]Examples are Ryu (1982); Hofstetter and Dozier (1986); Goidel and Langley (1995); Farnsworth and Lichter (2011); Garz (2014); Soroka (2016); Soroka et al. (2018); Bleich and van der Veen (2021); Kayser and Peress (2021) and Soroka and Krupnikov (2021).

[6]For example Soroka et al. (2019); Trussler and Soroka (2014); Leung and Strumpf (2022) and Dertwinkel-Kalt et al. (2022).

[7]Note that most of the studies however focus on a *negativity* bias and some conceptually don't even allow for a simultaneous positivity bias. If a positivity bias is found it tends to be smaller than the negativity distortion in most of the cases.

[8]We know from the information provision literature that economic decisions are based on beliefs people form through information they receive, and that such belief updating is sensible to framing effects (Haaland et al., forthcoming). It has been shown that emotional framing can impact a wide range of economically relevant outcomes such as tax compliance (Fišar et al., 2022), crime perceptions (Mastrorocco and Minale, 2018) or social trust and helping behavior (Han et al., 2019).

[9]This literature documents that the sentiment of news can impact investment decisions in the stock market,

how the sentiment of headlines translates into changes in readers' emotions and expectations.

The second relevant stream of literature discusses how the internet is changing the media landscape and media consumption patterns[10]. Most of the debate has so far focused on political dimensions such as the question of whether the internet amplifies the formation of filter bubbles and echo chambers. This paper is, to the best of my knowledge, the first to consider the effect of digitization on aggregate headline sentiment.

The paper closest to the one at hand is a field experiment by Balbuzanov et al. (2019) in which the authors experimentally vary payment schemes of writers in an online news firm in Kenya. They find that a pay-per-click contract substantially increases page views of the articles produced and changes the topics journalists choose to write about. My findings complement their field results with survey evidence and differ in two dimensions: First, my work emphasizes effects of pay-per-click on headline tonality, an outcome that Balbuzanov et al. (2019) do not consider. Second, my survey experiments allow for more control and a more fine-grained measurement, including the reaction of readers to the documented changes.

The remainder of the paper is organized as follows. Section 2 describes the descriptive analysis of media content. Section 3 illustrates the experiment with the journalists. Section 4 discusses how the results from the previous sections translate into shifts in economic behavior of readers. In Section 5 I regard all different results jointly, discuss their meaning and validity and conclude.

## 2. Is there a difference between online and offline headlines?

In this section I empirically compare the sentiment of headlines in the online and offline versions of media outlets. As a matter of fact, any detected difference in this section is correlational and not necessarily causal. However, to the best of my knowledge, the existence of different tonality in online and offline headlines has not yet been analyzed and thus such a comparison is an interesting first step to answer the research question at hand. After a plain comparison I consider the influence of the outlet-, time- and topic-fixed effects, the role of article length, agency content and controls for the tonality of the article content to better understand whether the observed differences are driven by or more pronounced in certain contexts.

### 2.1. Data

The data on the headlines of newspapers' online and offline versions are obtained through the data provider *LexisNexis*, which collects news articles, headlines and metadata on a daily basis. For the main analysis I regard headlines on economic issues of the German news outlets *BILD*, *Der Spiegel*, *Die Welt*, *Die Zeit* and *Rheinische Post*. These are all newspapers from the German market for which the database has both online and offline versions of the articles available in a cleanly separated way. This set of newspapers includes papers with different political orientations, target groups and writing styles. For example, *Die Welt* is considered rather conservative, while *Der Spiegel* is often entitled to be a more liberal news outlet. *BILD* is a tabloid while the rest

leading short-term under- or over estimations of stocks (Tetlock, 2015).

[10]See for example Gentzkow and Shapiro (2011); Pariser (2012); Boxell et al. (2017); Farrell (2012) and Allcott et al. (2020).

Table 1: Description of Datasets

**Data for main analyses: Only economic news, includes article content**

| news outlet | time-frame | N online | N offline | N total |
|---|---|---|---|---|
| *BILD* | 01/01/2017 - 01/06/2022 | 4,680 | 3,092 | 7,772 |
| *Der Spiegel* | 01/02/2003 - 01/06/2022 | 79,192 | 14,296 | 93,488 |
| *Die Welt* | 05/07/2009 - 01/06/2022 | 65,226 | 24,535 | 89,761 |
| *Die Zeit* | 10/01/2009 - 01/06/2022 | 50,643 | 64,260 | 114,903 |
| *Rheinische Post* | 10/09/2020 - 01/06/2022 | 11,348 | 22,593 | 33,941 |
| all from above | all available points in time | 211,089 | 128,776 | 339,865 |

**Data for robustness checks: Includes all topics, but headlines only**

| news outlet | time-frame | N online | N offline | N total |
|---|---|---|---|---|
| *BILD* | 01/01/2017 - 17/05/2022 | 66,720 | 97,576 | 164,296 |
| *Der Spiegel* | 01/01/2021 - 31/12/2021 | 19,122 | 4,896 | 24,018 |
| *The New York Times* | 01/01/2021 - 09/11/2021 | 25,230 | 29,216 | 54,446 |

*Notes:* Table 1 describes the timeframes and number of observations available for each news outlet that is part of the descriptive analyses. The data in for the main analyses contains articles on economic topics only, but comes with the entire content of the article. The datasets for the robustness checks news on all topics, but are limited to the headlines only.

of the newspapers are not. *Rheinische Post* targets a regional audience, the other for outlets publish news for the entire country.

The only restriction that was put on the articles before extracting them from the database was that they should be classified as talking about economic topics[11]. Thus, the entire time-frame for which both online and offline articles were available for a certain outlet is included. The headlines come together with information on their publication date, whether the headline was published online or offline, article length (scaled in 1000 words) and the full text of the article. Non-editorial content such as obituaries, letters to the editor or ads were dropped from the dataset.

It is noteworthy that the licensed time-frame differs substantially between outlets, which limits the time-frame for which I have data on all outlets to late 2020 until mid 2022. Table 1 provides an overview about the different time-frames and number of articles available for each news outlet. For the main analysis the data on *all* outlets is available is regarded. In Appendix A.5.1 I show that the main finding of this section reproduces for most outlets separately and independently of the time-frame regarded.

To assess the generalizability of my results I additionally look at headlines on non-economic issues in *BILD*, *Der Spiegel* and *The New York Times* in the robustness subsection. These datasets consists of *all* online and offline headlines of the outlets that were published in the denoted time-frames in the lower part of Table 1. That means that the content of these headlines is not restricted to economic news. In these robustness datasets the full articles are not included, and thus no sentiment classification or other text classification methods on the article level can be performed. The data however comes with information on the ressort[12] a specific article was published in and, for the offline versions, with the page number of the article.

[11] *LexisNexis* provides this classification as a standard service for all available news articles, the exact name of the classified group is "Economic News and Economic Indicators".

[12] Ressorts indicate the topic of the article in a relatively wide sense, such as "sports", "business" or "culture".

## 2.2. Empirical Strategy

I first classify the sentiment of the headlines, which generates my main outcome variables. These are then used in regressions to compare the intensity of sentimental language of the respective online and offline versions of the news outlets. Summary statistics of all variables used for the descriptive analyses are available in Table A2.

### 2.2.1. Sentiment Classification

There are multiple ways to classify text as negative, positive or neutral and the literature has not settled to one standard or best classifier. Instead, current research suggests that (i) the choice of the classifier can play a crucial role for outcomes and (ii) different text types require different classifiers[13] (Shapiro et al., 2022; Hartmann et al., 2022).

To assess which algorithm is the best one for the classification of my data use a random subset of the headlines that was classified by a research assistant and compare the fit of the algorithms' classifications with these human-coded sentiment categories. In particular I compare the popular lexicon-based approaches of the SentimentWortschatz (Remus et al., 2010), the LoughranMcDonald-dictonary (Loughran and McDonald, 2011) and the Valence Aware Dictionary and sEntiment Reasoner (Hutto and Gilbert, 2014) to a pre-trained version of the transformer-based machine learning model RoBERTa (Liu et al., 2019) and a version of the same machine learning model that I fine-tuned with another subset of classifications from a research assistant.[14]

Descriptions of all different classifiers are available in Appendix A.1. A comparison of the different approaches reveals that the fine-tuned Financial-RoBERTa outperforms all other models. I therefore use the classifications of this model for all further analyses. The accuracy and F1 scores of the different classifiers on my test set are available in Appendix A.1.3 in Table A1. As an robustness-check I show in Appendix A.5.3 that the main findings of this section reproduce when using any of the other considered algorithms.

### 2.2.2. Comparison

The two main outcomes I compare are *sentiment* and *emotionality*. For *sentiment* a positive headline is assigned the value "1", a neutral headline the value "0" and a negative headline the value "$-1$".[15] However, when looking at aggregate tonality, positive and negative headlines can cancel each other out with this approach. For example, a completely neutral news outlet would receive the same average sentiment score as a news outlet that produces 50% positive and 50% negative headlines and no neutral headlines at all. Therefore, I also compute *emotionality*, which is the absolute value of sentiment. Thus, sentiment gives an indication of the average direction of the tonality and emotionaity an indication of whether a headline is classified as neutral or not.

I compare sentiment and emotionality of the headlines online and offline using linear regressions. To facilitate the comparison of effect sizes in I standardize both emotionality and sentiment to

[13] Dictionaries created for social media data might for example miss-classify a lot of text from the financial domain and vice-versa.

[14] As only one of these algorithms is available for German data I follow recommendations of Shapiro et al. (2022) and translate my dataset with the *Google Translate API* prior to classification.

[15] This is a standard way to express the overall tonality of textual data.

have a mean of zero and a standard deviation of one.[16] Standard errors are bootstrapped based on clusters on the level of the news outlet[17]. Equation 1 describes the baseline specification, which is a simple online VS. offline comparison.

$$tonality_i = \beta_0 + \beta_1 online_i + \epsilon_i \quad (1)$$

$tonality_i$ is the outcome: either *sentiment* or *emotionality* of the headlines in standard deviations. $online_i$ is a dummy indicating whether the headline was published online or offline.

If there is any difference between the headlines, it could for example stem from journalists reporting more intensively online on days when emotional things happen. It could also be that journalists just report more about emotional topics online. Another possibility is that the effect might be driven by specific outlets or topics in the dataset. To explore whether this is the case I subsequently add control variables for outlet-, topic- and time-fixed effects as well as a measure of the tonality of the content to the baseline specification. I also run a regression in which I control for all available variables jointly. The regression equations and detailed descriptions of the variables used in these additional analyses are available in Appendix A.2.

## 2.3. Results

**Emotonality** As depicted in the histogram in Figure 1b the largest share of the headlines on economic issues are classified as neutral. When comparing the online headlines to the offline versions it becomes clear that headlines from the internet versions of the outlets are more often classified as emotional (=positive or negative).

A comparison of the means in the emotionality measure reveals that an online headline is 23

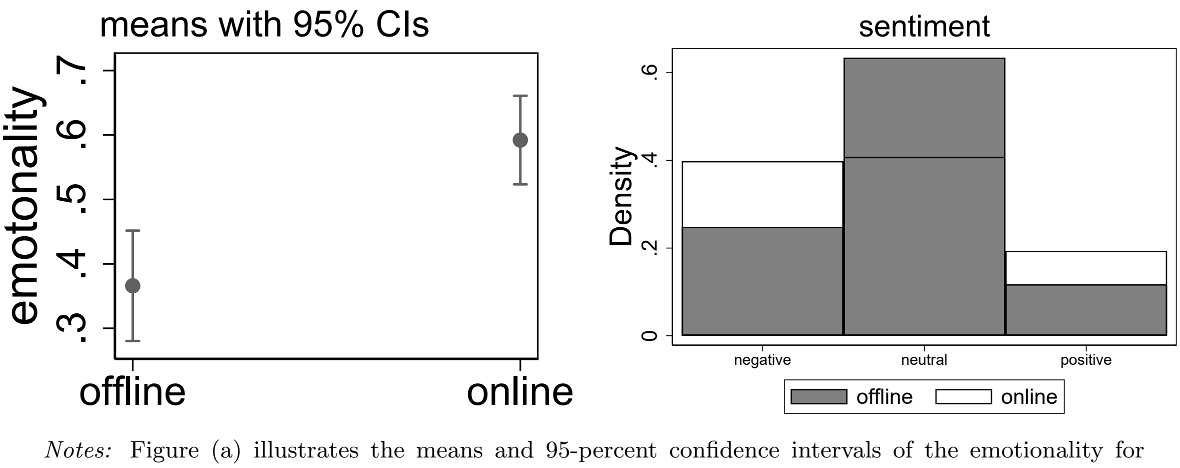
[16]This standardization takes away the interpretation in terms of probability of the emotionality outcome. I therefore repeat all describes analyses with the original measures in Appendix A.4.

[17]I use bootstrapping as the number of clusters (which is 5) is too small for standard clustering methods.

Figure 1: Descriptive Results: Emotionality and Sentiment of Headlines

(a) Means and 95%-confidence intervals

(b) Distribution




*Notes:* Figure (a) illustrates the means and 95-percent confidence intervals of the emotionality for the headlines online and offline. The estimates displayed here are obtained by running the baseline regression as described in equation 1 (no controls). Figure (b) depicts the distribution of the sentiment measure. Values for the offline headlines are shaded in gray.

percent more likely to be emotional than an offline headline ($p < 0.001$). This corresponds to a difference of 0.42 standard deviations. The respective means with 95-percent confidence intervals are depicted in Figure 1a.

The difference remains but becomes somewhat smaller when controlling for content tonality, time- and topic-fixed-effects as well as article length and agency content. In particular, controlling for the emotionality of the content reduced the difference in emotionality between online and offline headlines to 16 percent (0.31 standard deviations, $p < 0.001$). Controlling for topic-fixed-effects reduces the estimate to 18 percent (0.36 standard deviations, $p < 0.001$) and the time controls to 20 percent (0.39 standard deviations, $p < 0.001$).

Thus, all of these variables seem to play a role for the emotionality difference between online and offline headlines, but none is able to explain it entirely. Controlling for all mentioned variables jointly as well as adding controls for article length and agency content shrinks the difference to 13 percent (0.25 standard deviations, $p < 0.001$). Table 2 presents the results of the regression analyses with and without control variables.

**Sentiment** Regarding sentiment the data reveal that headlines online are on average more negative than offline. However, the sizes of the differences are substantially smaller than for the emotionality measure and lie, depending on how many controls are added, between 0.06 and 0.1 standard deviations. This smaller difference is probably at least partially due to the way the sentiment variable is constructed - the higher share of more positive and more negative headlines cancels out to some extent when comparing the sentiment means. Table 3 presents the results of the regression analyses with and without control variables. The difference in sentiment between online and offline headlines becomes smaller when controlling for content sentiment and topic fixed effects, but is not much affected by the other controls.

**Robustness** I assess the robustness of the main findings in three dimensions. First, the question arises whether the observed differences are driven by specific outlets or certain time periods. To assess this I separately regard the sentiment and emotionality of all considered outlets over time - which reveals that for almost all outlets online headlines are for all points in time more emotional than offline headlines. The picture for sentiment is more nuanced. While online headlines are on average more negative for all of the outlets regarded, this is not true for all outlets at all points in time. A detailed analysis of this is available in Appendix A.5.1. Second, I regard different sentiment classification algorithms. Appendix A.5.3 shows that the main findings reproduce with all different classification algorithms that were considered in section 2.2.1. Third, I assess whether my results are specific to German economic news. This does not seem to be the case, as my findings reproduce in a sample of non-economic headlines and in headlines from the *New York Times*. Details on these robustness checks are available in Appendix A.5.4.

Overall, the most robust finding of the descriptive analysis is that online headlines are on average more often emotional than offline headlines. Also, online headlines are on average more negative, but this may differ for specific outlets at specific points in time.

**Result 1: Headlines of news outlets are written more often emotionally for online audiences than for offline audiences.**

Table 2: OLS Estimates - Emotionality of Headlines in Standard Deviations

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| online | 0.4524*** | 0.3107*** | 0.3685*** | 0.3923*** | 0.2544*** |
| | (0.0730) | (0.0440) | (0.0604) | (0.0595) | (0.0312) |
| content emotionality | | 0.7206*** | | | 0.6570*** |
| | | (0.0353) | | | (0.0330) |
| article length | | | | | -0.0047 |
| | | | | | (0.0171) |
| agency content | | | | | 0.0726* |
| | | | | | (0.0376) |
| topic FE | no | no | yes | no | yes |
| time FE | no | no | no | yes | yes |
| Constant | -0.2809*** | -0.6064*** | -0.2580*** | 0.000 | -0.0145 |
| | (0.0874) | (0.0266) | (0.0919) | (0.0516) | (0.0168) |
| $R^2$ | 0.0482 | 0.1705 | 0.0889 | 0.0262 | 0.1667 |
| Observations | 339,865 | 339,865 | 339,865 | 339,865 | 339,865 |

*Notes:* Table 2 reports OLS estimates with bootstrapped standard errors in parentheses. Bootstrapping is conducted with 50 replications based on 5 clusters at the level of the news outlet. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.2. The offline headlines are always the reference group and the difference is expressed in standard deviations. Control variables are a content emotionality dummy, article length in 1000 words, a agency content dummy, outlet fixed effects as dummies for the different outlets, topics and points in time. The same regressions with the independent variable expressed as emotionality dummy are available in Appendix A.4. *** p<0.01, ** p<0.05, * p<0.1

Table 3: OLS Estimates - Sentiment of Headlines in Standard Deviations

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| online | -0.1065*** | -0.0599*** | -0.1078*** | -0.0878*** | -0.0622*** |
| | (0.0373) | (0.0139) | (0.0287) | (0.0245) | (0.0176) |
| content sentiment | | 0.7488*** | | | 0.7371*** |
| | | (0.0435) | | | (0.0409) |
| article length | | | | | -0.0362** |
| | | | | | (0.0177) |
| agency content | | | | | -0.0043 |
| | | | | | (0.0222) |
| topic FE | no | no | yes | no | yes |
| time FE | no | no | no | yes | yes |
| Constant | 0.0661 | 0.2540*** | 0.0692 | 0.0000 | 0.0143 |
| | (0.0406) | (0.0241) | (0.0606) | (0.0108) | (0.0128) |
| $R^2$ | 0.0027 | 0.2768 | 0.0188 | 0.0013 | 0.2765 |
| Observations | 339,865 | 339,865 | 339,865 | 339,865 | 339,865 |

*Notes:* Table 3 reports OLS estimates with bootstrapped standard errors in parentheses. Bootstrapping is conducted with 50 replications based on 5 clusters at the level of the news outlet. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.2. The offline headlines are always the reference group and the difference is expressed in standard deviations. Control variables are a content emotionality dummy, article length in 1000 words, a agency content dummy, outlet fixed effects as dummies for the different outlets, topics and points in time. The same regressions with the independent variable expressed as sentiment ranging from -1 to 1 are available in Appendix A.4. *** p<0.01, ** p<0.05, * p<0.1

## 3. What causes the difference?

This part of the paper investigates whether the documented differences in online and offline headlines indeed stem from stronger incentives for writers to generate clicks. An experiment with professional journalists provides a clear causal estimate for this relationship. Afterwards I explore a mechanism which might explain *why* there are different incentive structures in the first place: Increased competition on the headline level in the online sphere. For this I provide quasi-experimental and correlational evidence from case studies using the data from section 2.

### 3.1. Experiment with Professional Journalists

To understand better why the difference in headline valence observed in section 2 exists I conduct an online experiment with N=201 professional journalists. The data was collected between 17th of November 2021 and 14th of December 2021 in cooperation with the biggest German journalist association *DJV*. Participation was limited to full-time journalists. The survey was send to members of the association via E-Mail and they could start it anytime in the data collection period. The experiment was implemented with the survey software *Qualtrics*. The median time to complete it was 6.3 minutes. Participants were paid either according to click-rates of headlines they selected or with a flat rate. The average compensation was €7.73 and payments were made via PayPal. In addition to the experiment with journalists a survey experiment with a sample of N=299 "readers" was conducted. More details on this second experiment can be found in section 4. Ethical approval for the experiments was obtained by the Ethics Committee of the Faculty of Management, Economics and Social Sciences at University of Cologne (reference: 210036LM) and they were pre-registered in the AEA Social Science Registry as AEARCTR-0008658[18].

#### 3.1.1. Setting

Journalists had to choose a headline for a given, real article about an economic forecast for the German economy. For the interpretations of their decisions and the readers reactions it can be helpful to keep the economic situation in Germany at this point in time in mind: In November and December 2021 Germany was living through its fourth COVID-wave and the national vaccination campaign was faltering. The German stock index had just marked a new all-time-high above 16.000 points. Analysts were expecting the pandemic and possible new lock-downs to harm the economy, but they did not expect a harsh decline. Media reports were often emphasizing that the industries that were most affected by lock-downs were not central for the German GDP growth[19]. Most analysts expected the economy to continue growing, however on a relatively low rate[20].

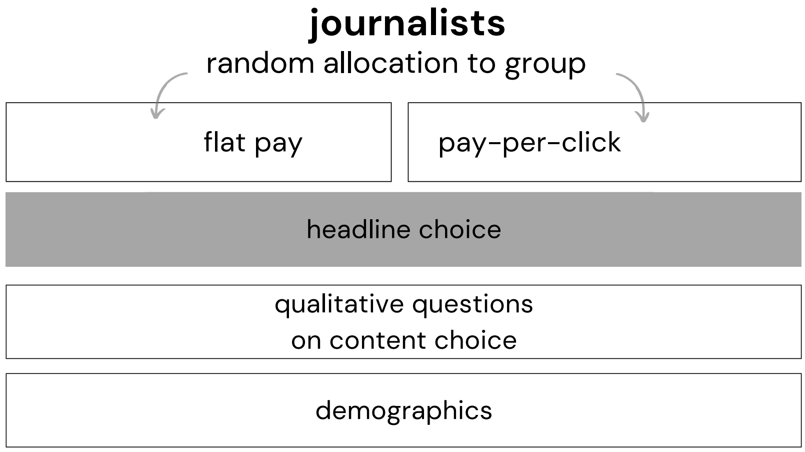
#### 3.1.2. Experimental Design

The experimental setup consists of two separate experiments: One with journalists and one with the readers. The experiments are independent of each other except for the payment of the treatment group of journalists. Their compensation is determined by the click-rates of the readers. I describe the experimental procedures of the experiment with journalists in the next

[18]You can find the pre-registration here.
[19]One example is a report from *Tagesschau*.
[20]See for example reports from *Die Zeit* and *Süddeutsche Zeitung*.

paragraphs and of the experiment with readers in section 4. Figure 2 provides an overview of the journalists' experiment.

Figure 2: Experimental Procedures - Overview




*Notes:* Figure 2 gives an overview of the experimental procedures. Randomization is indicated with arrows. The main outcome is shaded in grey.

In the journalist-experiment, a computer randomly allocates the participants into one of two groups: either the "flat pay" group or the "pay-per-click" group. Then, each journalist has to select a headline out of three suggestions for a given article. The article is a real report of the news agency *dpa* about the future development of the German economy[21]. The possible headlines are all factually correct, but emphasize different aspects of the topic so that one headline is positive, another one neutral and the last one negative[22]. The full wording of the article and the headlines as well as an translation to English are available in Appendix B.1.

The "flat-pay"-group is paid a flat rate for this task, while the remuneration of the treatment group depends on how often their selected headline is clicked on by readers. After they chose a headline journalists have the option to type a suggestion of a headline themselves as a free-text answer[23]. I also ask journalists to list factors they find relevant when it comes to choosing a headline for an article ("qualitative questions"). Finally, the following demographics and preferences are elicited: the journalists' seniority, age, political orientation and economic preferences (in a short-form of the survey modules suggested by Falk et al. (2018)). A translation of the full experimental instructions is available in Appendix D.1.

### 3.1.3. Sample Descriptives

The journalists sample consists of 201 subjects who are aged between 18 and 80 and have between two and 42 years of working-experience in the news industry. Compared to the overall German population, political preferences in the sample are distorted towards politically left-leaning parties,

[21] I chose an article from a news agency as those reports are usually brief and relatively neutral. They are not published by the agency, but a service for newsrooms who then in turn often adopt these reports slightly and publish them with their own modifications. Adoptions of the report I used in the experiment have for example been published by *Tagesschau*, *Zeit Online* or *Focus Online*, among others.

[22] All algorithms considered in section 2 classify the positive headline as positive, the neutral one as neutral and the negative one as negative.

[23] Note that this part of the experiment is not incentivized.

which is in line with previous findings in surveys among journalists. The distribution of the set of covariates as well as a comparison of the political preferences with the overall population is available in Appendix B.2.

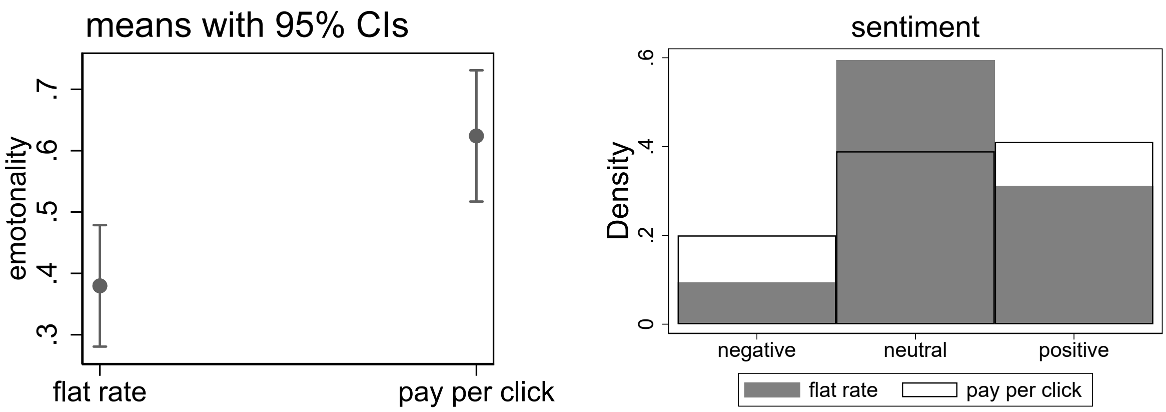
The randomization of the journalists into treatment and control group worked well. Out of the 20 considered covariates only one is statistically significantly higher in the control group than in the treatment group, as should be expected by chance.[24] The balance table for this sample is available in Appendix B.2 in Table B6.

#### 3.1.4. Results

**Headline choice** Figure 3a depicts the means of the headline choices of the journalists in the respective groups. Journalists who are paid per click are on average 20 percent more likely to choose an emotional headline than the journalists who received a flat payment (this corresponds to 0.41 standard deviations, p= 0.001). As can be seen by the distribution in Figure 3b this effect is driven by pay-per-click journalists selecting both the negative and the positive headline more often than the flat-pay group.

Figure 3: Headline Choice by Journalists

(a) emotionality of choice by group (b) distribution of choice




*Notes:* Figure 3a illustrates means in the "flat rate" and the "pay per click" group in terms of emotionality. The emotionality measure is 0 if a journalist selected the neutral headline and equals 1 if the journalist selected the positive or the negative headline. Figure 3b provides an overview of the distribution of the headline choices. The choices of the "flat rate" group are shaded in gray.

Table 4 presents the main regression analysis of the journalist experiment. In column (1) and (2) the outcome is a sentiment measure which equals 1 if the positive headline is chosen, 0 if the neutral headline is chosen and -1 if the negative headline is chosen. In column (3) and (4) the outcome is a dummy for headline emotionality which equals 1 if the negative or positive headline is chosen and 0 otherwise. The reference group is always the flat pay group. Column (1) and (3) are univariate linear regressions. In column (2) and (4) the set of covariates is added.

Journalists continue to be 20 percent more likely to choose an emotional headline when controlling for observables and the coefficients for this effect are statistically significant on the one

[24]In particular, participants in the flat pay group state statistically significantly more often to vote for other parties than those currently represented in the German parliament. Controlling for this imbalance does not change results.

Table 4: OLS Estimates - ATE on Headline Choice

| | Sentiment | | Emotions | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| pay-per-click | -0.0064548 | 0.0058872 | 0.2048659*** | 0.2082093*** |
| | (0.0971431) | (0.1000453) | (0.0694661) | (0.070733) |
| controls | no | yes | no | yes |
| Constant | 0.2169811*** | -0.7394713 | 0.4056604*** | -0.6645239 |
| | (0.0584536) | (0.5248686) | (0.0479311) | (0.4230708) |
| $R^2$ | 0.0000 | 0.1280 | 0.0418 | 0.1555 |
| Observations | 201 | 201 | 201 | 201 |

*Notes:* Table 4 reports OLS estimates with robust standard errors in parentheses. The flat rate group is the reference group. Control variables are age, seniority, role in the news organization, political orientation, education, phone-use, risk preference, patience, altruism, trust and narcissism. A table displaying all covariates in detail is available in Appendix B.3 as Table B7. *** p<0.01, ** p<0.05, * p<0.1

percent level. In terms of sentiment, I do not detect a statistically significant effect. The point estimates are furthermore very small and point in different directions depending on whether controls are added or not. The reason for this "null-effect" is that the positive and negative choices in the sentiment measure, on average, cancel out. A table with coefficients for all covariates is available in Appendix B.3 as Table B7.

**Result 2: Incentives to generate clicks induce journalists to select emotional headlines.**

**Free text headline** The text data from the free headline answer is analyzed using the sentiment algorithm from section 2 and hand-coding by a research assistant. None of these methods reveals any significant differences between the headline choices of the two groups. This might be the case for two reasons: First, this part of the experiment has not been monetarily incentivized and one could thus only expect differences in the answers if there were "spillovers" from the incentivation of the previous task or if participants had not understood the incentivation well. Second, a free text field comes with lots of degrees of freedom. If one expects variation as in the headlines analyzed in section 2 I would have needed roughly 658 journalists in my sample to detect an statistically significant effect on the five percent significance level.[25]

**Decision factors** In the qualitative question journalists stated that the most important factors for deciding on the headline of an article in general are comprehensibility, factual correctness, length and that the headline sparks curiosity. There is no statistically significant difference between the flat-pay and pay-per-click group in which factors the journalists regarded as relevant for headline decisions in general. An overview over all answers regarding these potential factors is available in Appendix B.4 in Figure B11.

[25] This number is the minimum required sample size to detect an effect of the size of the coefficient when I control for all considered covariates es in equation 5. As in the experiment not only the topic, sentiment and length were fixed, but the article was identical, the required sample size coming out of this power analysis is likely to be still a lower-bound estimate.

## 3.2. Mechanism

The experiment presented in the previous section shows that higher incentives to capture attention can induce journalists to select more emotional headlines. But *why* might these incentives be higher in online markets in the first place? A plausible reason is higher competition on the headline level. As stated in the motivation of this paper: While in offline markets, at most front-page headlines have to be catchy to increase sales, every headline competes with thousands of others in digital markets. In this subsection I provide anecdotal evidence that, consistent with this idea, supports a relationship between the degree of competition and headline tonality.

### 3.2.1. Headline Competition in Offline Markets

If it is true that in offline markets, at most front-page headlines have to be catchy to increase sales, then there should also be a difference in the tonality of front-page and non-front-page headlines in offline markets. To analyze this, I use the offline data of my robustness datasets and compare the emotionality and sentiment of front-page to non-front-page headlines.[26] This reveals that front-page headlines are indeed on average more often emotional and more often negative than headlines on other pages. The size of the difference ranges from the headlines being on average 0.07 to 0.26 standard deviations more emotional and 0.02 to 0.18 standard deviations more negative, depending on the considered outlet (all p-values are smaller than 0.01). Detailed results are available in Appendix B.5.1.

### 3.2.2. Headline Competition in Online Markets

To further explore the relationship of the degree of competition and headline tonality I copy the idea behind the identification of Meyer et al. (2022) and analyze a variation in the competition of headlines in online markets induced by a legal dispute by news companies and news aggregators in Germany in 2013. This dispute lead to the removal of one of the news outlets in my sample from popular news aggregators, while the other two remained[27]. Thus, headline competition in the online markets was reduced for one outlet and not for the others.

This setting allows for two comparisons: First, I compare the average tonality of the online headlines in the affected outlet to the other outlets before and after the removal. Then, I compare the average headline tonality of the online version of the removed outlet to the tonality of the offline headlines of the same outlet before and after the removal.

In line with the hypothesis that higher competition is related to higher emotionality, these comparisons suggest that being removed from the aggregators reduced the average emotionality of the online headlines of the *Die Welt*. This result is robust to the use of different sentiment classifiers and the consideration of different time-frames. I do not detect any evidence for effects in terms of sentiment. Details on this analysis are available in Appendix B.5.2.

[26]The reason that I don't use the offline data of my main dataset is that this data does not come with the page numbers of the articles.

[27]I have only data on three outlets (Der Spiegel, Die Welt and Die Zeit) as this time-frame is not covered in my data for BILD and Rheinische Post.

# 4. Do emotional headlines impact economic outcomes?

The previous sections have shown that headlines online are more emotional than offline and that this difference can at least partially be explained by shifts in incentives for journalists. But how should this development be evaluated? Is it good or bad, important or negligible? To be better able to interpret the described amplification of news sentiment it can be helpful to understand whether it changes economic outcomes in any way. This is what this section investigates. I use the second experiment (the one with the readers) to explore if emotional headlines can have economic impacts on emotions, beliefs and investment decisions of their readers.

The survey experiment with a sample of N=299 participants was conducted in the same time-period as the journalists-experiment. Readers were recruited from the subject pool of the Cologne Laboratory for Economic Research (CLER) via ORSEE (Greiner, 2015). It took them a median of 4.7 minutes to complete the survey and they received on average €1,98 as compensation. Again, payments were made via PayPal. This second experiment was used for the elicitation of the click-rates in order to pay the journalistic treatment-group, but also to obtain some estimates of how readers react to the different headlines they were exposed to.

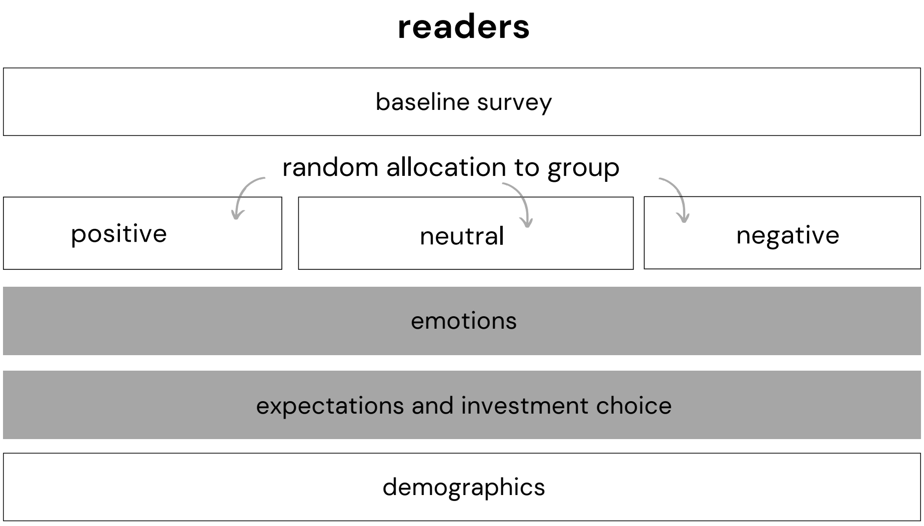
## 4.1. Experimental Design

Readers start the experiment with a baseline survey that elicits their current feelings, risk preferences and a self-evaluating question about their knowledge in economics and finance. Then, a computer randomly exposes them to one of the three headlines from the journalists' experiment.

The readers have the option to click on the headline to read the entire article. Such a click costs them €0,05 of their remuneration (which equals around five percent of their entire compensation).[28] Readers are told that the content of the article might help them to make more

[28]This is done for generating a cost of clicking so that not everyone chooses to click on the headline.

Figure 4: Experimental Procedures - Overview




*Notes:* Figure 4 gives an overview of the experimental procedures. Randomization is indicated with arrows. The main outcomes are shaded in grey.

informed decisions in payoff-relevant questions at a later stage of the experiment. The clicking decisions of the readers here determine the payment of the "pay-per-click" group of journalists.

After reading the headline and taking a clicking-decision, the readers' current feelings as well as beliefs about the content of the article (=the future development of the German economy) are elicited. Feelings are measured with a general question about the current mood (11 point Likert scale) and with the i-PANAS-sf scale (Thompson, 2007), which is a popular scale to measure short-term emotions in psychological studies[29].

Beliefs are elicited in an incentivized, numerical forecasting task regarding the GDP growth and the DAX development. The incentivation of these measures works as follows: For each variable, one reader is randomly selected and paid depending on the accuracy of his or her expectations (with up to €10).[30] Each reader can receive at most one payoff for one of the expectations. This incentive scheme comes with the properties that subjects are not able to hedge risk between expectations. Further, the setting is non-strategic (i.e. the expected payoff is independent from the expectations of the other subjects). This incentivation should thus induce readers to provide answers that reflect their true expectations.

Also, readers have to make a decision in an incentivized investment task. They can decide which amount of money (between 0 and 50 cent) to invest in the DAX until the end of 2021 and until the end of 2022. The rest of the money is payed out at the same date as the investment is payed out. A summary of the experimental procedures is illustrated in Figure 4. A translation of the full experimental instructions for the readers is available in Appendix D.2.

### 4.2. Sample Descriptives

Readers were recruited from the subject pool of the Cologne Laboratory for Economic Research and thus consist mainly of students and recent graduates. 44% of the participants are male and the median age is 27. In terms of covariates, the sample is mostly balanced. Out of the 45 t-tests conducted as a balance test 5 differences between the treatment groups are statistically significant different from zero on conventional significance levels. These differences do not seem to be systematic for one group and results do not change when controls for observables are added. The balance table for this sample is available in Appendix B.2 as Table C15.

### 4.3. Results

I find evidence for reactions of readers in terms of their short-term emotions and expectations, but don't detect any statistically significant reactions in investment decisions and clicking behavior. Participants seem to react differently to the headlines depending on their gender.

#### 4.3.1. Emotional reactions

**Current mood** Participants who read the positive article state to feel significantly better than those who read the neutral or negative headlines in the general current mood question (0.42

[29]The i-PANAS-sf scale measures affect in the ten dimensions. The positive dimensions are if participants feel active, inspired, determined, attentive and alert. The negative dimensions are whether they feel afraid, nervous, ashamed, hostile and upset. Those ten dimensions are each elicited on a 5-point Likert scale and combined in one index by summing up the positive dimensions and subtracting the negative ones.

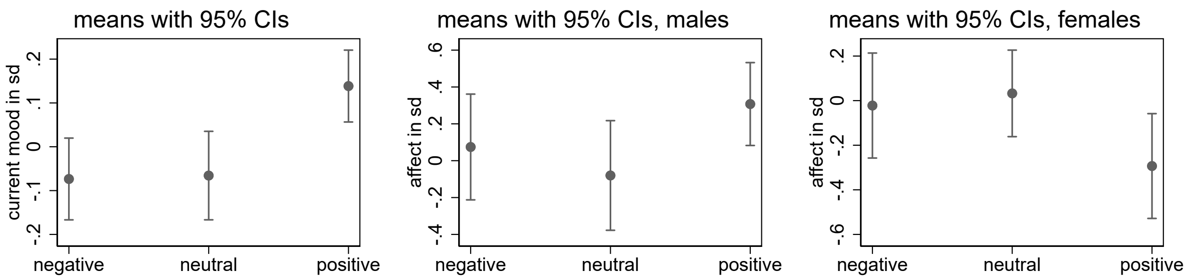
[30]I did not disclose the exact payment formula in more detail to subjects.

points difference on the 11-point Likert scale, p=0.001). If I normalize the outcome in standard deviations this corresponds to a 0.2 standard deviations higher mood. This result is illustrated in Figure 5a. A table displaying regressions for both the current mood questions and the affect scale is available in Appendix C.2 as Table C16.

**Affect** Perhaps surprisingly, I do not find evidence on effects of the different headlines on aggregate affect. This is due to a heterogeneous reaction to the headlines by gender as depicted in Figures 5b and 5c. Participants who identify as male feel on average more determined when they read the negative headline and more attentive and less afraid and upset when they read the positive one (always relative to the neutral condition). Thus, the positive headline seems to have an activating and positive effect on male's affect. Participants identifying as female instead feel on average less active if they are exposed to the negative or positive headline. Further, when they read the positive headline, they state to be less determined and alert. This results in an aggregate negative effect of the positive headline on affect and thus the results on males and females cancel out when they are regarded simultaneously. This heterogeneous reactions in terms of gender are discussed in greater detail in section 4.4.

Figure 5: Emotional Reactions by Readers

(a) average mood (b) average affect among males (c) average affect among females




*Notes:* Figure 5a illustrates means of the mood in the different treatment groups expressed in standard deviations. Figures 5b and 5c provide the means of aggregate affect in standard deviations by gender.

### 4.3.2. Expectations

**Directly related expectations** The article participants are exposed to talks about forecasts of German GDP growth, and thus their expectations about future German GDP growth are the most direct reflection of how participants use the information they receive in the article to update their beliefs. I compute the distance to the forecast stated in the article by using the absolute value of subtracting the belief of the actual forecast. As participants are able to insert any value they want this outcome is prone to outliers. Thus, the answers are winsorized[31] and expressed in standard deviations to facilitate interpretation.

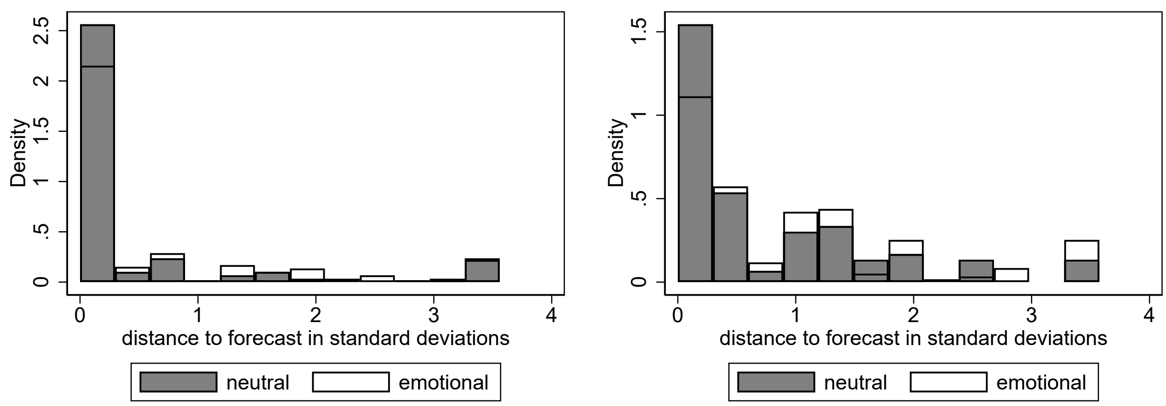
The data shows that the distance to the forecast is higher for GDP expectations of those who read an emotional headline, but only the estimate for 2022 is statistically significant on conventional levels. There, the expectations of participants who read an emotional headline

[31]For the main specification I winsorize outcomes at the 95th percentile, but I obtain similar results if I use the 99th percentile instead.

were on average 0.23 standard deviations (p=0.047) further away from the forecast than the expectations of those who read the neutral headline. Figure 6 illustrates the distribution of the distances to the forecasts. Regression results are available in Appendix C.2.2 in Table C17.

Figure 6: Belief Updating by Readers

(a) 2021 (b) 2022




*Notes:* Figure 6 illustrates distribution of the distance to the forecasts expressed in standard deviations. The answers of the positive and negative group are pooled as "emotional" for better legibility. The answers of the neutral group are shaded in gray.

**Indirectly related expectations** To understand whether readers use the headlines to extrapolate on expectations in related, but not directly mentioned topics, I also elicit the readers' expectations for the future development of the German stock index DAX. As this index is compiled of the 40 biggest German companies it is related to German GDP growth, but changes in the German growth rate do not perfectly translate into changes in the DAX. To anchor the estimates I provide all readers with official data for the value of the DAX from the week prior to the experiment. Again, the answers are winsorized and standardized.

I find that readers on average lower their expectations for the DAX development when they are exposed to the negative headline by 0.27 standard deviations for 2021 (p=0.043) and by 0.29 standard deviations for 2022 (p=0.039). This corresponds to a change in expectations of 148 to 356 basis points, which is roughly the magnitude of average daily volatility of the DAX. I only find evidence for reactions to the positive headline for males in 2022. They expect the DAX to perform 0.38 standard deviations *worse* if they are exposed to the positive headline (p=0.043), while females, if anything, expect it to develop better. Regression results on the DAX expectations are available in Appendix C.2.2 as table C18.

**Investment decisions** I find no evidence for effects of the different headlines on decision making in the incentivized investment decision task.

> **Result 3: The emotionality of headlines can translate into alterations in readers' emotions and belief updating.**

**Click rates** There is no statistically significant difference on the click rates of the headline, except for the difference between the positive and the neutral headline which is significant at the ten percent level (t test, p=0.0626). This difference vanishes when controlling for the standard set of covariates in a regression analysis.

### 4.4. Heterogeneous effects by gender

Readers react differently to the headlines depending on their gender. For example, the distortions in the belief updating of the GDP expectations and the more negative DAX expectations after the negative headline are mainly driven by female participants. The only evidence for reactions in expectations of male participants is on the positive headline on DAX expectations - they expect the DAX to perform worse in 2022 when they read the positive headline[32].

A possible explanation for these heterogeneous reactions may lie in the topic that I chose for the experiments. It could be that male and female participants hold on average different attitudes towards the economy and stock markets ex-ante. In line with this idea, recent evidence by Henkel and Zimpelmann (2022) shows that women tend to view the stock market more negative than men on average. If this is also the case in my sample and if my participants exhibit confirmation bias, it could be that only the female participants react strongly to the negative headline because it confirms their priors much more than those of male participants. The positive headline in turn provokes negative affect in women (in particular feeling less active, determined and alert), which might indicate a mismatch with their priors and thus provoke no update in beliefs. Overall, the heterogeneity in these results suggests that it could be interesting to test reactions of readers to emotional headlines in a broader setting with more than one topic.

### 4.5. Limitations

The reactions of readers to the headlines provide some insights into potential economic effects of more emotional news headlines. However, it is possible that this part of the experiment is under-powered and results should thus be interpreted with caution. One problem for power is that not everyone clicks on the article and decides to read it, thus detected effects on the distance to the value stated in the article are to be interpreted as intention-to-treat effects. True effects might be higher and smaller effects might just not be detected. Because clicking on the headline might be endogenous, I don't include it as a control variable in my regressions. Also, it cannot be instrumented for (with assignment to treatment as an instrument), because assignment to treatment is not relevant for explaining clicking behavior[33].

Because of the given limitations I plan to run an additional experiment with a larger sample that is representative for the German population as readers and elicit their reactions to emotional and neutral headlines on a wider range of topics.

[32] A possible explanation for this on the first glance surprising reaction is that participants might think that the market currently overreacts to the positive news and that there might be a recovery of this overreaction in 2022.

[33] Note that the "null finding" on the clicking rates does not imply that headline tonality is not important for clicks or that journalists are bad at predicting readers' behaviors. Instead, the clicking task in this experiment resembles more an information-seeking task and is not very representative of how readers take clicking decisions in the online sphere. Studies that look at actual clicking behavior and headline tonality in the field find that emotionality, and especially negativity of headlines is indeed related to higher clicking rates.

## 5. Discussion and Conclusion

In this paper, I provide evidence for an amplification of the sensationalist slant of news headlines due to the digitization of media markets. I demonstrate that (i) headlines of a wide range of news outlets are more emotional online than offline, (ii) a reason for this are the incentives to generate clicks, and (iii) more emotional headlines can provoke changes in emotions and economic expectations of their readers. I thereby improve our understanding of how digital media markets shape economic outcomes. As the majority of news is already consumed online, such a difference to analogue media markets is highly relevant and will probably become even more important in the future.

The descriptive and experimental parts of this paper come with the typical limitations for these kinds of analyses, but they mitigate each other to some degree if the results are regarded jointly. For instance, the main limitation of the descriptive part is that any documented difference is correlational, and not necessarily causal. However, this part of the paper can give us confidence in the external validity, as it demonstrates that a difference in the tone of headlines online and offline is widespread in the real world. On the flip side, the main limitation of the experimental part is its external validity, and its key strength is the clear identification of a causal effect. What further improves the credibility of the experimental part is the non-standard, and very realistic subject pool of professional journalists. Therefore, section 2 and 3 taken together imply that news headlines are more emotional online than offline and a reason for this is the incentive to generate clicks. Section 3.2 provides evidence for a plausible reason for why incentives to capture attention might be higher online in the first place: increased competition of headlines in online markets.

Online and offline news environments however differ in many more dimensions than just the monetary clicking incentive as modeled in the experiment. The reader should therefore keep in mind that the experimentally demonstrated effect is likely an important, but probably not the only factor that drives the difference in emotional slant online and offline. Journalists in the real world might for example have an incentive to generate emotional headlines for the online sphere even when they are not monetarily incentivized, as they might get direct utility simply out of gaining a lot of attention for their work. Also, in contrast to analogue markets there are almost no space limits in the digital sphere, which makes the use of computer-automated articles that target a very niche audience feasible. In Appendix B.5.3 I show that the introduction of such "robo-news" can explain a kink in the positivity of headlines in one of the outlets in my sample.

Further, caution should be taken when comparing the effect sizes of the descriptive and the experimental part to understand how much of the difference in observed emotionality is driven by stronger incentives to generate clicks. While the size of the effect in the experiment (0.41 standard deviations) seems to be similar to the difference in the descriptive data (0.45 standard deviations), such an exercise is comparing apples to oranges, because the journalists in the experiment were exposed to much more structure than in the real word. In particular, they had to choose one headline out of three suggestions instead of writing it themselves. The actual effect of clicking incentives in an environment without that much control could be smaller or larger. I therefore recommend to interpret my results mainly qualitatively and to not focus too much on the exact sizes of the coefficients.

The experiment with the readers suggests that emotional headlines can distort emotions and economic expectations, which may cause them to take sub-optimal decisions. In this way my findings suggest that incentives of journalists play an important role for information transmission from news to readers and thereby the quality of news. This might be interesting from an policy perspective, for example when policymakers set the incentive schemes for journalists in public service media organizations. It seems quite likely that such policymakers face a trade-off: They can reach more people with emotional headlines, but the information these people receive will be biased. The paper at hand is only able to shed light on one side of this problem: The distortions in beliefs and emotions due to more emotional headlines. Investigating this trade-off from both sides and analyzing how it could optimally be solved seems to be a promising avenue for future research.

## References


Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow (2020) "The Welfare Effects of Social Media," *American Economic Review*, **110** (3), 629–76.

Arango-Kure, Maria, Marcel Garz, and Armin Rott (2014) "Bad News Sells: The Demand for News Magazines and the Tone of Their Covers," *Journal of Media Economics*, **27** (4), 199–214.

Balbuzanov, Ivan, Jared Gars, and Emilia Tjernström (2019) "Media and motivation: the effect of performance pay on writers and content."

Bleich, Erik and A. Maurits van der Veen (2021) "Media portrayals of Muslims: a comparative sentiment analysis of American newspapers, 1996–2015," *Politics, Groups, and Identities*, **9** (1), 20–39.

Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro (2017) "Greater Internet use is not associated with faster growth in political polarization among US demographic groups," *Proceedings of the National Academy of Sciences*, **114** (40), 10612–10617.

Dertwinkel-Kalt, Markus, Johannes Münster, and Dainis Zegners (2022) "If it Bleeds, it Leads: Negativity in Online News," *Working Paper*.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde (2018) "Global Evidence on Economic Preferences*," *The Quarterly Journal of Economics*, **133** (4), 1645–1692.

Farnsworth, S. and S. R. Lichter (2011) "The nightly news nightmare," *Lanham MD: Rowman Littlefield*.

Farrell, Henry (2012) "The Consequences of the Internet for Politics," *Annual Review of Political Science*, **15** (1), 35–52.

Fišar, Miloš, Tommaso Reggiani, Fabio Sabatini, and Jiří Špalek (2022) "Media negativity bias and tax compliance: experimental evidence," *International Tax and Public Finance*, **29**, 1160–1212.

Garz, Marcel (2014) "Good news and bad news: evidence of media bias in unemployment reports," *Public Choice*, **161**, 499–515.

Gentzkow, Matthew and Jesse M. Shapiro (2011) " Ideological Segregation Online and Offline *," *The Quarterly Journal of Economics*, **126** (4), 1799–1839.

Goidel, Robert K. and Ronald E. Langley (1995) "Media Coverage of the Economy and Aggregate Economic Evaluations: Uncovering Evidence of Indirect Media Effects," *Political Research Quarterly*, **48** (2), 313–328.

Greiner, Ben (2015) "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, **1** (1), 114–125.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart (forthcoming) "Designing Information Provision Experiments," *Journal of Economic Literature*.

Han, Lei, Rui Sun, Fengqiang Gao, Yuci Zhou, and Min Jou (2019) "The effect of negative energy news on social trust and helping behavior," *Computers in Human Behavior*, **92**, 128–138.

Hartmann, Jochen, Mark Heitmann, Christian Siebert, and Christina Schamp (2022) "More than a Feeling: Accuracy and Application of Sentiment Analysis," *International Journal of Research in Marketing*.

Henkel, Luca and Christian Zimpelmann (2022) "Proud to Not Own Stocks: How Identity Shapes Financial Decisions," *ECONtribute Discussion Paper No. 206*.

Hofstetter, C. Richard and David M. Dozier (1986) "Useful News, Sensational News: Quality, Sensationalism and Local TV News," *Journalism Mass Communication Quarterly*, **64** (4), 815–53.

Hutto, C. J. and E. E. Gilbert (2014) "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text."

Kayser, Mark A. and Michael Peress (2021) "Does the Media Cover the Economy Accurately? An Analysis of Sixteen Developed Democracies," *Quarterly Journal of Political Science*, **16** (1), 1–33.

Klamm, Christopher, Ines Rehbein, and Simone Ponzetto (2022) "FrameASt: A Framework for Second-level Agenda Setting in Parliamentary Debates through the Lense of Comparative Agenda Topics," *ParlaCLARIN III at LREC2022*.

Leung, Tin Cheuk and Koleman S Strumpf (2022) "All the Headlines that Are Fit to Change," *SSRN*.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov (2019) "RoBERTa: A Robustly Optimized BERT Pretraining Approach."

Loughran, Tim and Bill McDonald (2011) "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *The Journal of Finance*, **66** (1), 35–65.

Mastrorocco, Nicola and Luigi Minale (2018) "News media and crime perceptions: Evidence from a natural experiment," *Journal of Public Economics*, **165**, 230–255.

Meyer, Tim, Anna Kerkhof, Carmelo Cennamo, and Tobias Kretschmer (2022) "Competition for Attention on Information Platforms: The case of local news outlets," *Working Paper*.

Pariser, Eli (2012) *The Filter Bubble: How the New Personalized Web is Changing what We Read and how We Think*: Penguin Books.

Remus, R., G. Heyer, and U. Quasthoff (2010) "SentiWS - a Publicly Available German-language Resource for Sentiment Analysis."

Rozin, Paul and Edward B. Royzman (2001) "Negativity Bias, Negativity Dominance, and Contagion," *Personality and Social Psychology Review*, **5** (4), 296–320.

Ryu, Jung S. (1982) "Public Affairs and Sensationalism in Local TV News Programs," *Journalism Mass Communication Quarterly*, **59** (1), 74–137.

Shapiro, Adam Hale, Moritz Sudhof, and Daniel J. Wilson (2022) "Measuring news sentiment," *Journal of Econometrics*, **228** (2), 221–243.

Soleimanian, Mohammad (2022) "Do Firms Walk Their Talk in Corporate Social Responsibility Reports? – Evidence From Forward-Looking Statements."

Soroka, Stuart (2016) "Gatekeeping and the Negativity Bias," *Oxford Research Encyclopedia of Politics*.

Soroka, Stuart, Mark Daku, Dan Hiaeshutter-Rice, Lauren Guggenheim, and Josh Pasek (2018) "Negativity and Positivity Biases in Economic News Coverage: Traditional Versus Social Media," *Communication Research*, **45** (7), 1078–1098.

Soroka, Stuart, Patrick Fournier, and Lilach Nir (2019) "Cross-national evidence of a negativity bias in psychophysiological reactions to news," *Proceedings of the National Academy of Sciences*, **116** (38), 18888–18892.

Soroka, Stuart and Yanna Krupnikov (2021) *The Increasing Viability of Good News*, Elements in Politics and Communication: Cambridge University Press.

Tetlock, Paul C. (2015) "Chapter 18 - The Role of Media in Finance," in Simon P. Anderson, Joel Waldfogel, and David Strömberg eds. *Handbook of Media Economics*, **1** of Handbook of Media Economics: North-Holland, 701–721.

Thompson, Edmund R. (2007) "Development and Validation of an Internationally Reliable Short-Form of the Positive and Negative Affect Schedule (PANAS)," *Journal of Cross-Cultural Psychology*, **38** (2), 227–242.

Trussler, Marc and Stuart Soroka (2014) "Consumer Demand for Cynical and Negative News Frames," *The International Journal of Press/Politics*, **19** (3), 360–379.

# A. Supplementary Materials Descriptive Analysis

## A.1. Description of Sentiment Classifiers

The numerous classifiers available can be sorted into machine learning approaches and lexicon-based techniques. While the most recent machine learning classifiers oftentimes produce a higher accuracy (bench-marked with human-coded data), lexicon-based classifiers are much less of a black-box in their classification decisions.

### A.1.1. Dictonary-based approaches

For the dictonary-based classifications I remove stopwords from the headlines using stopword dictionaries and stem the headlines in a first step[34]. Then, the number of positive and negative words in each headline is counted. The *sentiment score* is then the number of positive words minus the number of negative words, which is a standard way to express the overall tonality of textual data with dictionary approaches. To enable a comparison with the human coded classifications I then define a headline to be positive if the sentiment score is larger than zero, negative when it is smaller than zero and neutral when it is equivalent to zero.

**SentimentWortschatz (SentiWS)**
The SentimentWortschatz (SentiWS) is a publicly available German language resource for sentiment analysis. With around 1,650 positive and 1,800 negative basic forms (which results in around 16,000 positive and 18,000 negative word forms including the various inflected forms) it is to the best of my knowledge the largest German-language sentiment dictionary.

**LoughranMcDonald-dictonary (LM)**
The LoughranMcDonald-dictonary (LM) is a widely used English-language finance-specific dictionary. As I consider news articles on economic issues this dictonary might be a better fit than general sentiment dictionaries.

**Valence Aware Dictionary and sEntiment Reasoner (VADER)**
A common mistake in sentiment classifications by dictionary approaches happens when valence shifters are not taken into account. For example the sentence “I am not happy” would be classified as positive, as it contains the word “happy”. This problem is addressed by the Valence Aware Dictionary and sEntiment Reasoner (VADER). This algorithm is a lexicon and rule-based sentiment analysis approach that in addition to word counts respects negations and other commonly used language rules. The likelihood of the above described mistakes is here therefore much lower, which could increase the accuracy of the classifications. VADER was however mainly created to classify social media content, which in turn might lower accuracy. The classifications of VADER are distributed on a scale from -2 (very negative) to +2 (very positive). Again, for the comparison with the human coded data I consider a headline to be positive if the VADER score is above zero, negative when it is smaller than zero and neutral when it is equivalent to zero.

### A.1.2. Machine learning approaches

**Financial-RoBERTa (pre-trained)**
The machine learning technique I use is called Robustly Optimized Bidirectional Encoder Representations from Transformers (RoBERTa). It is an optimized BERT pretraining approach which is currently often considered the state-of-the-art for text classifications (Shapiro et al.,

[34]Stopwords are words that usually do not contain any meaning such as “the” or “a”. They are removed to decrease computation time. For the English headlines I use the stopword-dictonary from the R tidytext package. For the German headlines I use a stopwords list from the countwordsfree blog.

2022). BERT itself is a self-supervised machine learning technique introduced by Google in 2018 for Transformer-based Natural Language Processing models. The model that I use is called Financial-RoBERTa and was pre-trained to analyze sentiment of financial texts. The training data included financial statements, earnings announcements, earnings call transcripts, corporate social responsibility reports, and news articles on environmental, social, governance and finance topics (Soleimanian, 2022). Financial-RoBERTa sorts the headlines into the three classes positive, negative or neutral which can be directly compared with the human-coded data.

**Financial-RoBERTa (fine-tuned)**
For further fine-tuning of Financial-RoBERTa I split the human-coded dataset (N=2500) into a training-, evaluation- and test-dataset[35]. The training data is used to train an additional layer of the Financial-RoBERTa model and the evaluation data is used to fine-tune the hyper-parameters of the training process. I use grid-search to fine tune the hyper-parameters and end up with a learning rate of 0.00002, 4 training epochs, a batch size per device during training of 16, a total number of 230 steps and 50 warm-up steps[36]. Again, this model sorts the headlines into the three classes which can be directly compared with the human-coded data.

### A.1.3. Evaluation of the classifications

I then use the test-dataset to evaluate and compare all of the classification of the different algorithms. Table A1 provides an overview of the accuracy and macro F1 scores of each model. In addition to the *sentiment* classifications it is interesting for my research question to know whether an algorithm is able to correctly identify emotional language independently of it being positive or negative. I therefore define an *emotonality* dummy which is equal to one if a headline was classified as positive or negative and zero otherwise and compute accuracy and F1 scores for this outcome as well.

Table A1: Evaluation of Sentiment Classification Algorithms

| | *Sentiment* | | *Emotionality* | |
|---|---|---|---|---|
| **algorithm** | **accuracy** | **macro F1** | **accuracy** | **F1** |
| SentiWS | 0.4773 | 0.3729 | 0.5413 | 0.3174 |
| LM | 0.6106 | 0.5257 | 0.6373 | 0.5436 |
| VADER | 0.5920 | 0.5554 | 0.6800 | 0.6428 |
| pre-trained roBERTa | 0.6853 | 0.6638 | 0.7200 | 0.7301 |
| fine-tuned roBERTa | 0.7253 | 0.7057 | 0.7413 | 0.7581 |

*Notes:* Table A1 presents the accuracy and the (macro) F1 scores for sentiment and emotionality of the different sentiment classifiers on the test-dataset. Accuracy is the share of correct classifications, while the F1-score is calculated considering both the precision and recall of classifications. The macro F1 averages the F1 score of the different classes (positive, neutral, negative).

[35] I follow the convention of using 70% of the data as training data, 15% as evaluation data and the remaining 15% as test-data. For the dataset at hand this results in 1750 training observations, 375 evaluation observations and 375 test observations.

[36] The seed was set to "1234".

## A.2. Comparisons with Controls

In addition to the baseline specification as described in equation 1, I run my descriptive analysis controlling subsequently for different potential drivers of the difference in tonality of online and offline headlines. The regressions that I run are described in Equations 2 to 4.

$$tonality_i = \beta_0 + \beta_1 online_i + \beta_2 content\_tonality_i + \epsilon_i \tag{2}$$

$$tonality_{ik} = \beta_0 + \beta_1 online_i + \beta_2 topic_k + \epsilon_{ik} \tag{3}$$

$$tonality_{it} = \beta_0 + \beta_1 online_i + \beta_2 time_t + \epsilon_{it} \tag{4}$$

$content\_tonality_i$ is a measure of sentiment or emotionality of the content of the article (not the headline). It is obtained by classifing the content with the same fine-tuned sentiment classifier as the headlines. For coherence this measure is equivalent to the contents' emotionality when I analyze emotionality of the headline as an outcome and equivalent to the contents' sentiment when the sentiment of the headline is regarded.

$time_t$ is a vector with dummies for each day in the data.

$topic_k$ is a categorical variable that contains the most prominent topic out of 21 possibilities. The considered topic categories are: International, Defense, Government, Civil Rights, Environment, Transportation, Law and Crime, Energy, Health, Domestic Commerce, Immigration, Labor, Macroeconomy, Agriculture, Social Welfare, Technology, Education, Housing, Foreign Trade, Culture, Public Lands. These categories are assigned to the headlines using the ParlBERT-Topic-German model (Klamm et al., 2022).

Equation 5 describes a specification where I use all described covariates jointly and additionally control for article length[37] and agency content[38].

$$\begin{aligned} tonality_{ikt} = \beta_0 + \beta_1 online_i + \beta_2 content\_tonality_i \\ + \beta_3 topic_k + \beta_4 time_t + \beta_5 length_i + \beta_6 agency_i + \epsilon_{ikt} \end{aligned} \tag{5}$$

$length_i$ is a numerical variable with the number of words in the article that belongs to the headline.

$agency_i$ is a dummy variable equal to 1 if a news agencies name is included in the author line and 0 otherwise.

In Equation 5 the $\beta_1$ thus corresponds to the difference in tonality between online and offline headlines independent of the tonality of the article, time of publication, topic, agency-content and length. If there is still a difference detectable an explanation could be that journalists frame their stories differently for the online sphere, i.e. emphasize the emotional aspects of a story more often in the headline.

[37] This control is added as longer reads might be edited differently than quickly produced stories.

[38] This control is added because the process of choosing a headline for a pre-written story by a news agency might be different from the one for an article that was written by a journalist of the publishing outlet.

## A.3. Data

Table A2: Summary statistics of dependent variables, entire dataset

| Variable | Mean | Std. Dev. | Min. | Max. | N |
|---|---|---|---|---|---|
| content emotionality | 0.573 | 0.494 | 0 | 1 | 339,865 |
| content sentiment | -0.289 | 0.699 | -1 | 1 | 339,865 |
| article length | 0.753 | 0.688 | 0.001 | 71.382 | 339,865 |
| agency content | 0.381 | 0.485 | 0 | 1 | 339,865 |
| *Outlet* | | | | | |
| Rheinische Post | 0.099 | 0.299 | 0 | 1 | 339,865 |
| BILD | 0.023 | 0.149 | 0 | 1 | 339,865 |
| Der Spiegel | 0.275 | 0.446 | 0 | 1 | 339,865 |
| Die Welt | 0.264 | 0.440 | 0 | 1 | 339,865 |
| Die Zeit | 0.338 | 0.473 | 0 | 1 | 339,865 |
| *Topic* | | | | | |
| Agriculture | 0.022 | 0.148 | 0 | 1 | 339,865 |
| Civil | 0.084 | 0.278 | 0 | 1 | 339,865 |
| Culture | 0.005 | 0.076 | 0 | 1 | 339,865 |
| Defense | 0.015 | 0.122 | 0 | 1 | 339,865 |
| Domestic | 0.178 | 0.382 | 0 | 1 | 339,865 |
| Education | 0.006 | 0.081 | 0 | 1 | 339,865 |
| Energy | 0.022 | 0.148 | 0 | 1 | 339,865 |
| Environment | 0.030 | 0.172 | 0 | 1 | 339,865 |
| Foreign | 0.010 | 0.101 | 0 | 1 | 339,865 |
| Government | 0.223 | 0.416 | 0 | 1 | 339,865 |
| Health | 0.026 | 0.159 | 0 | 1 | 339,865 |
| Housing | 0.015 | 0.124 | 0 | 1 | 339,865 |
| Immigration | 0.008 | 0.091 | 0 | 1 | 339,865 |
| International | 0.115 | 0.319 | 0 | 1 | 339,865 |
| Labor | 0.024 | 0.155 | 0 | 1 | 339,865 |
| Law | 0.033 | 0.180 | 0 | 1 | 339,865 |
| Macroeconomic | 0.069 | 0.253 | 0 | 1 | 339,865 |
| Public | 0.002 | 0.048 | 0 | 1 | 339,865 |
| Social Welfare | 0.019 | 0.139 | 0 | 1 | 339,865 |
| Technology | 0.037 | 0.188 | 0 | 1 | 339,865 |
| Transportation | 0.046 | 0.209 | 0 | 1 | 339,865 |
| *Time* | | | | | |
| 2022 | 0.068 | 0.252 | 0 | 1 | 339,865 |
| 2021 | 0.127 | 0.333 | 0 | 1 | 339,865 |
| 2020 | 0.106 | 0.308 | 0 | 1 | 339,865 |
| 2019 | 0.085 | 0.279 | 0 | 1 | 339,865 |
| 2018 | 0.070 | 0.255 | 0 | 1 | 339,865 |
| 2017 | 0.057 | .231 | 0 | 1 | 339,865 |
| 2016 | 0.046 | 0.210 | 0 | 1 | 339,865 |
| 2015 | 0.054 | 0.227 | 0 | 1 | 339,865 |
| 2014 | 0.053 | 0.225 | 0 | 1 | 339,865 |
| 2013 | 0.059 | 0.237 | 0 | 1 | 339,865 |
| 2012 | 0.053 | 0.224 | 0 | 1 | 339,865 |
| 2011 | 0.043 | 0.204 | 0 | 1 | 339,865 |
| 2010 | 0.044 | 0.206 | 0 | 1 | 339,865 |
| 2009 | 0.046 | 0.209 | 0 | 1 | 339,865 |
| 2008 | 0.018 | 0.134 | 0 | 1 | 339,865 |
| 2007 | 0.016 | 0.128 | 0 | 1 | 339,865 |
| 2006 | 0.013 | 0.115 | 0 | 1 | 339,865 |
| 2005 | 0.012 | 0.109 | 0 | 1 | 339,865 |
| 2004 | 0.011 | 0.106 | 0 | 1 | 339,865 |
| 2003 | 0.009 | 0.097 | 0 | 1 | 339,865 |

## A.4. Results

Table A3: OLS Estimates - Emotionality of Headlines

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| online | 0.2261*** | 0.1553*** | 0.1842*** | 0.1961*** | 0.1272*** |
| | (0.0365) | (0.0220) | (0.0302) | (0.0297) | (0.0156) |
| content emotionality | | 0.3602*** | | | 0.3285*** |
| | | (0.0176) | | | (0.0165) |
| article length | | | | | -0.0023 |
| | | | | | (0.0086) |
| agency content | | | | | 0.0363* |
| | | | | | (0.0188) |
| topic FE | no | no | yes | no | yes |
| time FE | no | no | no | yes | yes |
| Constant | 0.3659*** | 0.2032*** | 0.3774*** | 0.000 | -0.0072 |
| | (0.0437) | (0.0133) | (0.0459) | (0.0258) | (0.0084) |
| $R^2$ | 0.0482 | 0.1705 | 0.0889 | 0.0262 | 0.1667 |
| Observations | 339,865 | 339,865 | 339,865 | 339,865 | 339,865 |

*Notes:* Table 2 reports OLS estimates with bootstrapped standard errors in parentheses. Bootstrapping is conducted with 50 replications based on 5 clusters at the level of the news outlet. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.2. The offline headlines are always the reference group. Control variables are a content emotionality dummy, article length in 1000 words, a agency content dummy, outlet fixed effects as dummies for the different outlets, topics and points in time. *** p<0.01, ** p<0.05, * p<0.1

Table A4: OLS Estimates - Sentiment of Headlines

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| online | -0.0734*** | -0.0413*** | -0.0743*** | -0.0605*** | -0.0429*** |
| | (0.0257) | (0.0096) | (0.0197) | (0.0169) | (0.0121) |
| content sentiment | | 0.5162*** | | | 0.5081*** |
| | | (0.0015) | | | (0.0282) |
| article length | | | | | -0.0249** |
| | | | | | (0.0122) |
| agency content | | | | | -0.0029 |
| | | | | | (0.0153) |
| topic FE | no | no | yes | no | yes |
| time FE | no | no | no | yes | yes |
| Constant | -0.1311*** | -0.0016 | -0.1290*** | 0.000 | 0.0098 |
| | (0.0279) | (0.0166) | (0.0418) | (0.0074) | (0.0088) |
| $R^2$ | 0.0027 | 0.2768 | 0.0188 | 0.0013 | 0.2765 |
| Observations | 339,865 | 339,865 | 339,865 | 339,865 | 339,865 |

*Notes:* Table 3 reports OLS estimates with bootstrapped standard errors in parentheses. Bootstrapping is conducted with 50 replications based on 5 clusters at the level of the news outlet. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.2. The offline headlines are always the reference group. Control variables are a content emotionality dummy, article length in 1000 words, a agency content dummy, outlet fixed effects as dummies for the different outlets, topics and points in time. *** p<0.01, ** p<0.05, * p<0.1

## A.5. Robustness checks

In this section I assess the robustness of my findings with respect to different points in time, different sentiment classifiers and different news markets.

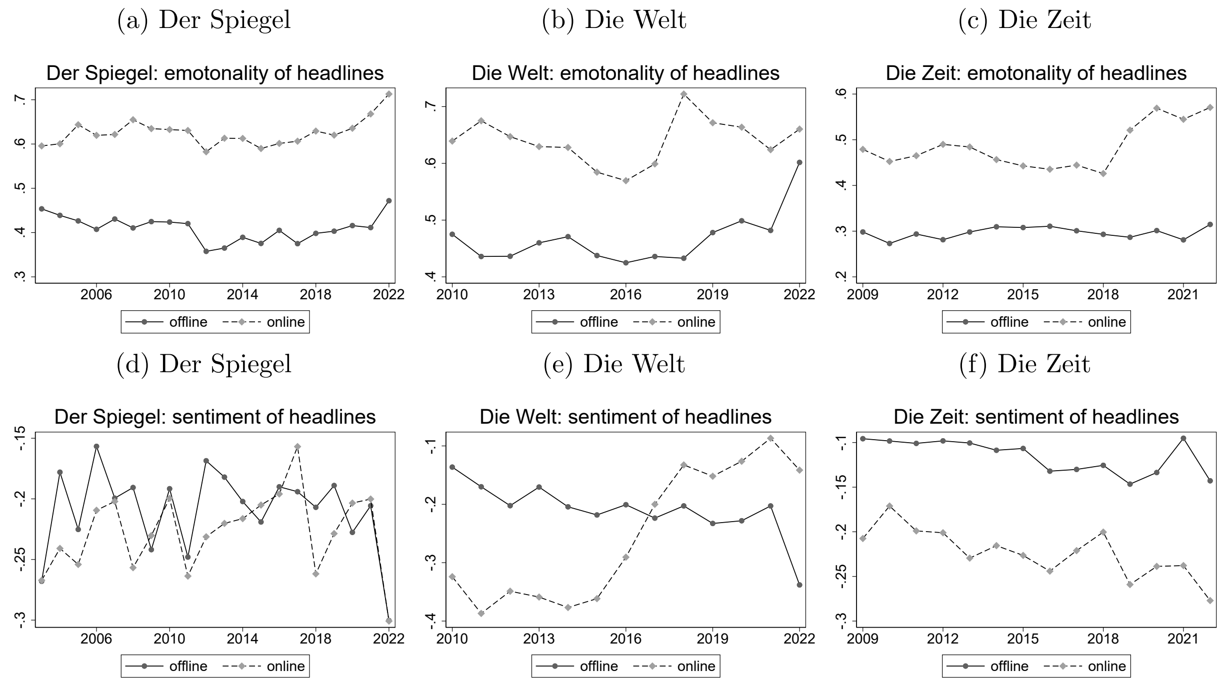
### A.5.1. Time Trends: Overall picture

Analyzing how the sentiment and emotionality of headlines develops over time is interesting because it can help to understand whether the differences in these measures are persistent or only present in certain periods. Further, to be able to interpret the results in a meaningful way it can be helpful to know whether the difference has been increasing, decreasing or remained fairly stable over the past years.

Figure A1 addresses these questions by plotting the development of both emotionality and sentiment over time for the three outlets for which the longest time-frame is available in my dataset. The difference in terms of emotionality seems to be quite stable. It persists in all years in all three outlets and no clear time-trend (in terms of increasing or decreasing emotionality) is visible across all outlets. The difference in sentiment however seems to be highly outlet-specific. For *Der Spiegel* it fluctuates a lot from year to year. While the online headlines are on average more negative for most years, there exist some years for which the reverse is true. For *Die Zeit* the difference seems to be quite persistent. The online headlines are in every year more negative than their offline counterparts. For *Die Welt* a similar picture existed from 2010 until 2015, but online headlines have become much more positive since 2016. They are by now even more positive than offline headlines on average. A more detailed analysis on this perhapt surprising shift is available in Appendix A.5.2.

Overall, the analysis of time trends reveals that the difference in emotionality seems to be fairly stable over time. The difference in sentiment in turn seems to be more specific for certain outlets or points in time.

Figure A1: Emotionality and Sentiment of Headlines over Time

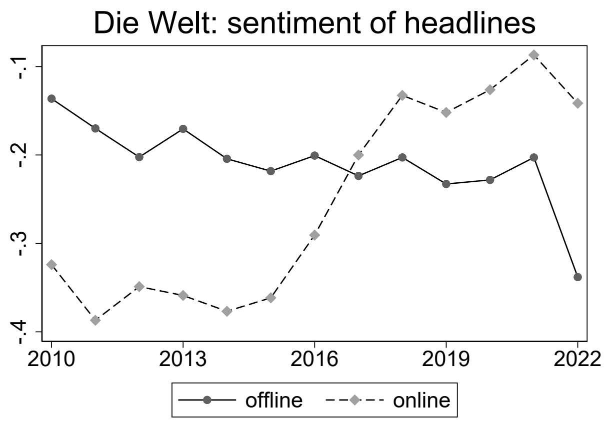



*Notes:* Figures A1(a) - A1(c) illustrate the development of emotionality over time. Figures A1(d) - A1(f) illustrate the development of sentiment over time. Values for the offline headlines are illustrated by a full line. The online values are depicted with a dotted line.

### A.5.2. Time trends: Shift in Sentiment for Die Welt

The analysis of time trends revealed the maybe surprising observation that while online headlines for *Welt Online* were more negative than their offline counterparts, this shifted in 2016 and by now online headlines are more positive for the outlet.

Figure A2: Sentiment of Welt Headlines over time

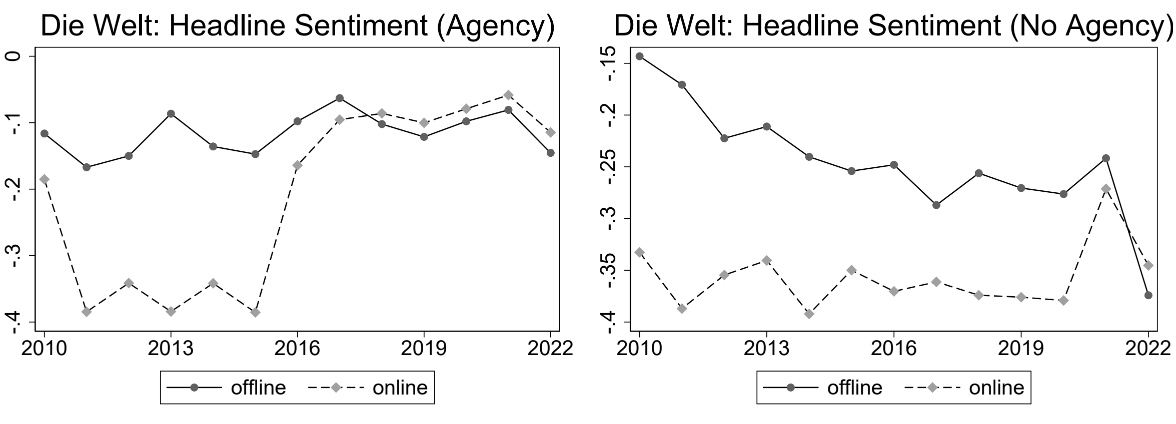



This trend is not observable for other outlets and thus raises the question why it occurred. A more detailed look into what drives the difference reveals that it is purely driven by a larger amount of agency content. Figure A3 displays the development of sentiment separately for agency and non-agency content.

Figure A3: Sentiment of Welt Headlines by Agency Content

(a) Agency Content

(b) No Agency Content




Looking at the absolute numbers of agency content articles published per month shows that simultaneous to the shift in sentiment the number of agency content articles published increased tremendously. From October 2015 until October 2016 the average number of agency content articles on economic issues every month is 42. For the same period one year later the average is 215. Table A5 provides an overview of the number of online agency content articles by *Die Welt* in my main dataset from 2015 until 2017.

I searched for changes on *welt.de* in this time-period using the internet archive WaybackMachine to understand what might cause this substantial increase in agency content. Interestingly, there was a shift in the *News Ticker* section on 13th of September 2016. Previously, this "ticker" was

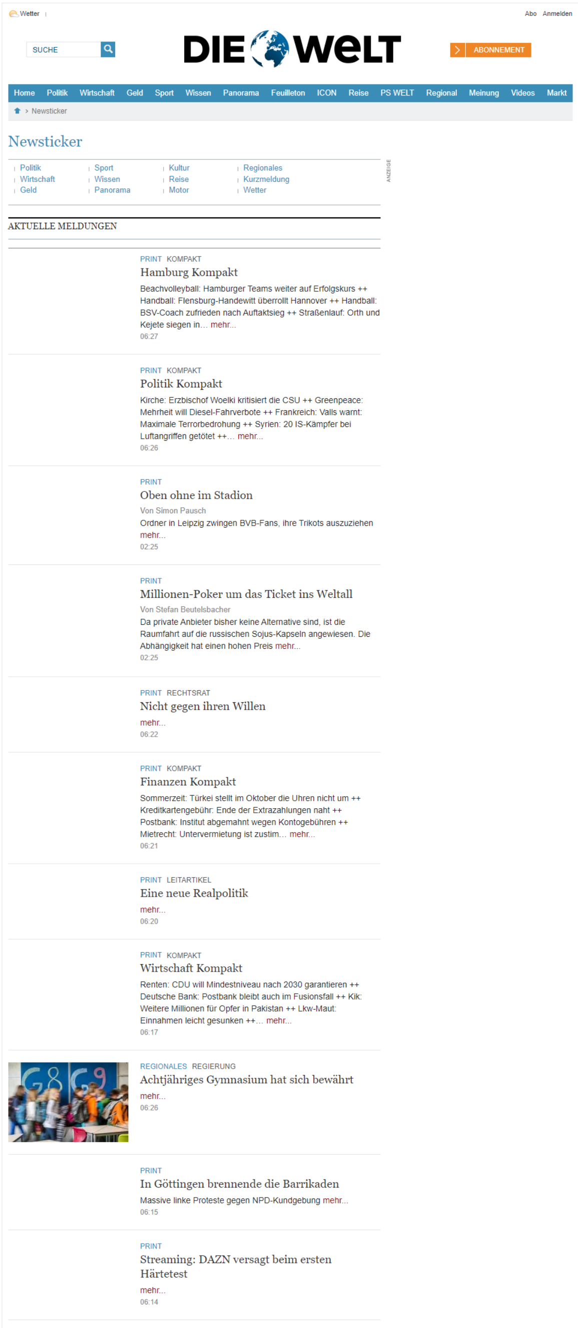
Table A5: Count of Online Agency Articles by *Die Welt* in Main Dataset

| **2015** | | **2016** | | **2017** | |
|---|---|---|---|---|---|
| *month* | *N* | *month* | *N* | *month* | *N* |
| January | 74 | January | 38 | January | 209 |
| February | 95 | February | 52 | February | 220 |
| March | 79 | March | 47 | March | 268 |
| April | 75 | April | 36 | April | 202 |
| May | 74 | May | 31 | May | 217 |
| June | 92 | June | 31 | June | 224 |
| July | 70 | July | 26 | July | 227 |
| August | 59 | August | 25 | August | 222 |
| September | 41 | September | 53 | September | 207 |
| October | 75 | October | 90 | October | 187 |
| November | 30 | November | 221 | November | 203 |
| December | 48 | December | 179 | December | 140 |

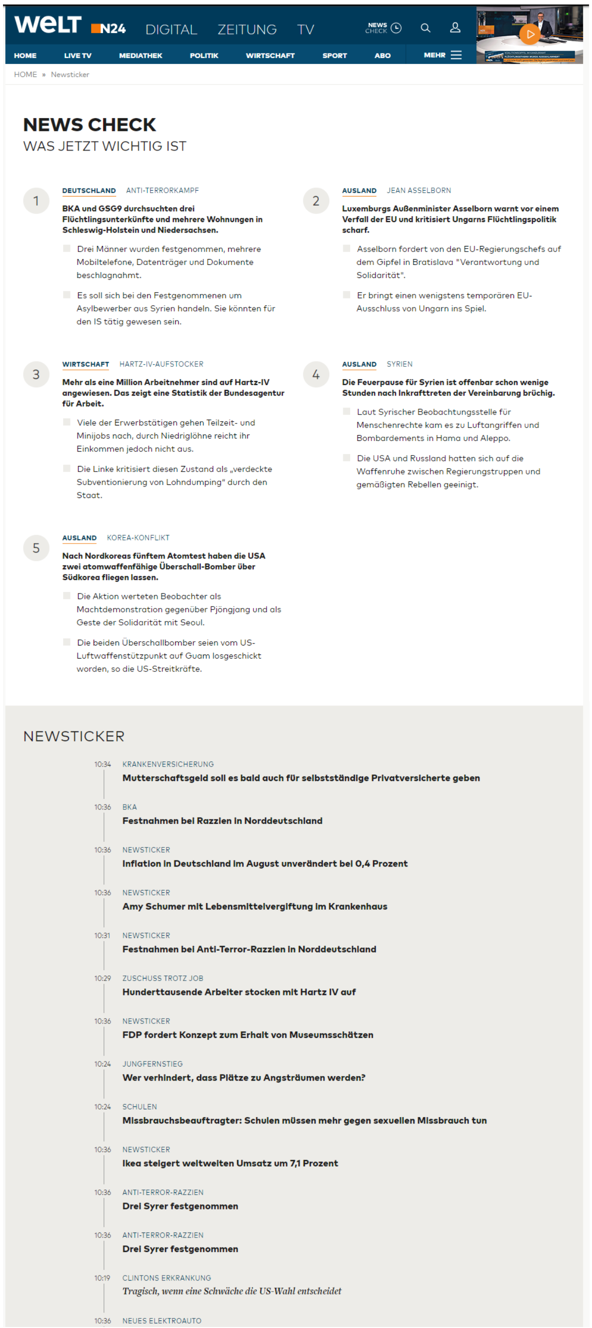
an overview page where most of the articles that used agency content were listed, but also a lot of other articles such as direct copies from the print version. Seemingly, it was not used frequently, as it still featured an outdated website design (the main landing page had already switched to a newer design at that point in time). From 13th of September onwards this ticker page also used the new website design and featured an area where the most important news were summarized as well as a news ticker in which a lot of agency content entered. It is therefore possible that the introduction of this new ticker came with an automatic publication of much more agency content, and that this in turn increased the share of the agency content in the sample to such an extend that it could shift the difference in average sentiment between online and offline headlines. Figure A4 contains screenshots of the ticker page before and after the change. The previous version is available via WaybackMachine here, the newer version here.

Figure A4: News Ticker at Welt Online

(a) Before 13.09.2016



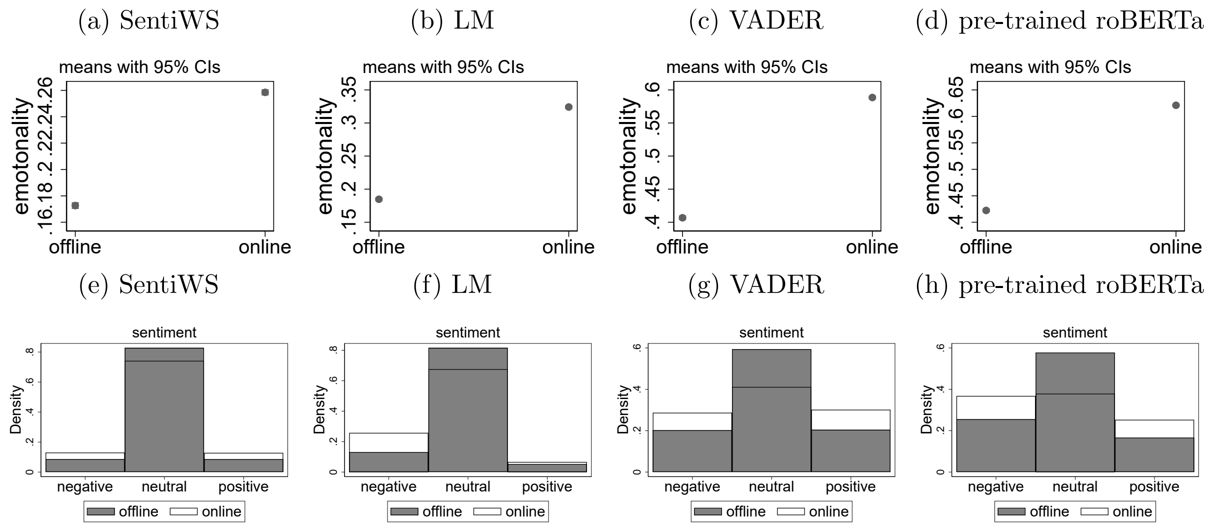
(b) After 13.09.2016



### A.5.3. Classification Methodology

Section A.1.3 demonstrates not only that the fine-tuned roBERTa model is the best of the considered for classifying the headlines in the dataset, but also that all of the classification approaches have some explanatory value[39]. They can thus be used to repeat the conducted analyses and assess to which degree the results are driven by the specific classification methodology.

Figure A5: Robustness Check: Classifications with Different Algorithms




*Notes:* Figures A5(a) - A5(d) illustrate the means and 95 percent confidence intervals in emotionality online and offline (without controls). Figures A5(e) - A5(h) depict the distribution of the sentiment variable with the different classification algorithms. The offline headlines are shaded in gray.

This robustness check reveals that the finding that online headlines are on average more emotional holds when using any of the considered algorithms. According to these estimations they are between 0.20 and 0.39 standard deviations more emotional than offline headlines. When all available controls are added (as in equation 5) the difference ranges from 0.14 to 0.27 standard deviations.

The finding that headlines online are on average more negative reproduces mostly, but not with all classification algorithms. Without controls the estimations range from headlines online being 0.02 standard deviations more positive to 0.23 standard deviations more negative than offline. When all control variables are added the difference ranges from headlines being 0.02 standard deviations more positive to 0.15 more negative online. Figure A5 illustrates the emotionality means and sentiment distributions of the respective classifications.

[39]By this I mean that they are better than random guessing.

### A.5.4. News Markets

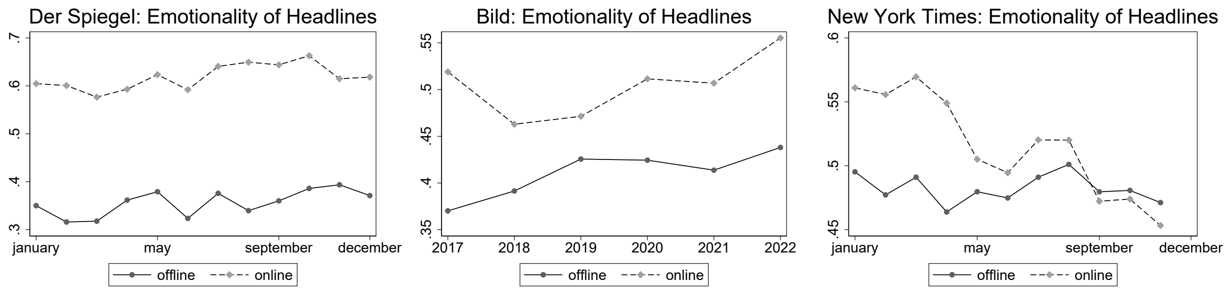
The main dataset consists only of articles on economic issues from German news outlets. Do the findings reproduce when articles on other topics or outlets from other countries are considered? To assess this I use additional datasets as described in the lower part of table 1. I consider headlines on all kinds of topics from *Der Spiegel*, *BILD* and *The New York Times* (but having a mostly much shorter time-frame available). As the content of the articles in not available in this data it is impossible to control for the article tonality in the subsequent regressions. Apart from this all estimations are equivalent to those described in section 2.2.2.

Comparing the estimates for *Der Spiegel* and *BILD* can help to understand how specific the findings are to articles on economic issues. The findings that headlines online are more emotional and negative than offline reproduce in these samples. Interestingly, the estimates are even bigger when compared to the estimate for the subset of economic articles for the same outlet. For example, online headlines from *Der Spiegel* are 0.52 standard deviations more emotional and 0.18 standard deviations more negative than offline (compared to 0.43 and 0.03 standard deviations in the economic article subset). A similar difference in effect size exists for *BILD*. This suggests that the estimations of the difference in sentiment of economic articles might be a relatively conservative estimate of the average difference for all topics.

The estimates might however be specific to the German news market. I therefore collect data on headlines of one of the most important American newspapers, the *New York Times*, and repeat my analysis with it. Again, the main findings so far reproduce: Headlines in the *New York Times* are online on average more emotional and negative than offline. The effect size for emotionality is with 0.08 standard deviations however substantially smaller than the average for the German market (0.45 standard deviations). Figure A6 provides an overview over these comparisons by illustrating the emotionality of headlines online and offline in these additional datasets.

Figure A6: Robustness Check: Classifications with Different Datasets

(a) Der Spiegel (b) BILD (c) The New York Times




*Notes:* Figures A6(a) - A6(c) illustrate the development of emotionality over time in the robustness datasets. The data from *Der Spiegel* and *The New York Times* are from 2021. Values for the offline headlines are illustrated by a full line. The online values are depicted with a dotted line.

# B. Supplementary Materials Causes of the Difference

## B.1. Journalists' experiment: article and headline

Figure B7: Given article in experiment - German

*"In ihrem im Oktober veröffentlichten Herbstgutachten gehen führende Wirtschaftsforschungsinstitute von einem Wachstum des Bruttoinlandsprodukts um 2,4 Prozent in 2021 aus. Im Frühjahr hatten sie noch damit gerechnet, dass in diesem Jahr ein Anstieg um 3,7 Prozent zu erwarten sei.*

*Die wirtschaftliche Lage in Deutschland sei nach wie vor von der Coronapandemie gekennzeichnet, hieß es. Im Verlauf des Jahres 2022 dürfte die deutsche Wirtschaft aber wieder die Normalauslastung erreichen. Laut Prognose der Institute steigt das Bruttoinlandsprodukt im Jahr 2022 um 4,8 Prozent. In ihrer Frühjahrsprognose gingen die Institute nur von einem Plus um 3,9 Prozent für das nächste Jahr aus."*

Quelle: dpa

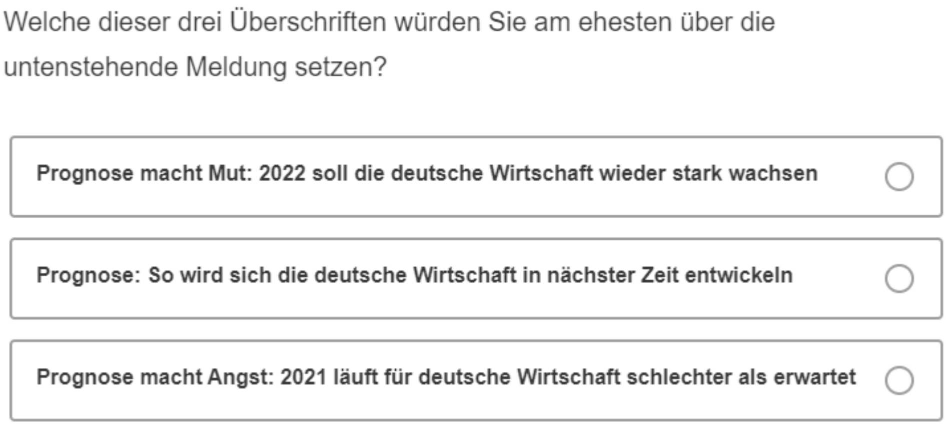
*Notes:* Figure B7 contains the text of the article as it was displayed to journalists and readers. A English translation can be found below. This article has originally been a short report of the German news agency *dpa*.

**English translation**

*In their autumn report, which published in October, leading economic research institutes expect gross domestic product to grow by 2.4 percent in 2021. In spring they had still expected an increase of 3.7 percent for this year.*

*The economic situation in Germany is still characterized by the corona pandemic, the report says. In the course of 2022, however, the German economy is expected to return to normal capacity. According to forecasts by the institutes, gross domestic product will increase by 4.8 percent in 2022. In their spring forecast, the institutes had only assumed an increase of 3.9 percent for the next year.*

Figure B8: Possible headlines for given article - German



Welche dieser drei Überschriften würden Sie am ehesten über die untenstehende Meldung setzen?

- Prognose macht Mut: 2022 soll die deutsche Wirtschaft wieder stark wachsen
- Prognose: So wird sich die deutsche Wirtschaft in nächster Zeit entwickeln
- Prognose macht Angst: 2021 läuft für deutsche Wirtschaft schlechter als erwartet

*Notes:* Figure B8 contains the original wording of the available headline choices. A translation to English can be found below. The headlines have been compiled by the researcher and evaluated by 10 professional journalists as being in principle realistic formulations for a headline.

**English translation**

Which of the following headlines would you most likely put above the article below?

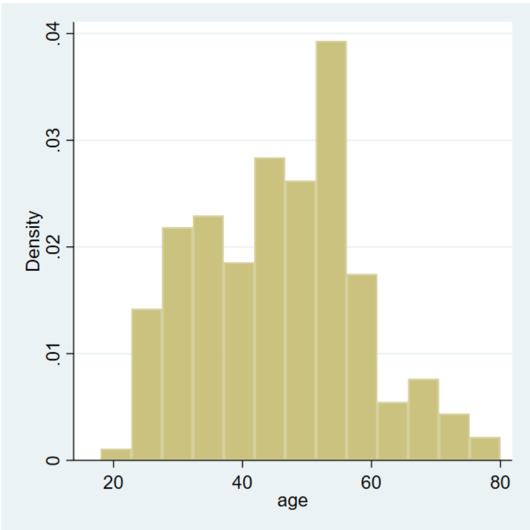
- Encouraging forecast: the German economy is expected to grow strongly again in 2022
- Forecast: This is how the German economy will develop in the near future
- Scary forecast: 2021 will be worse than expected for the German economy

Complete experimental instructions are available in Appendix D.1.

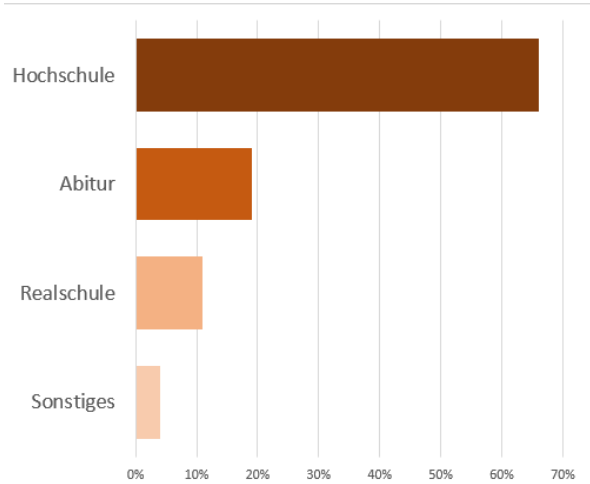
## B.2. Sample Characteristics and Randomization Check

Figure B9: Age and Education Distributions of Journalists

(a) age

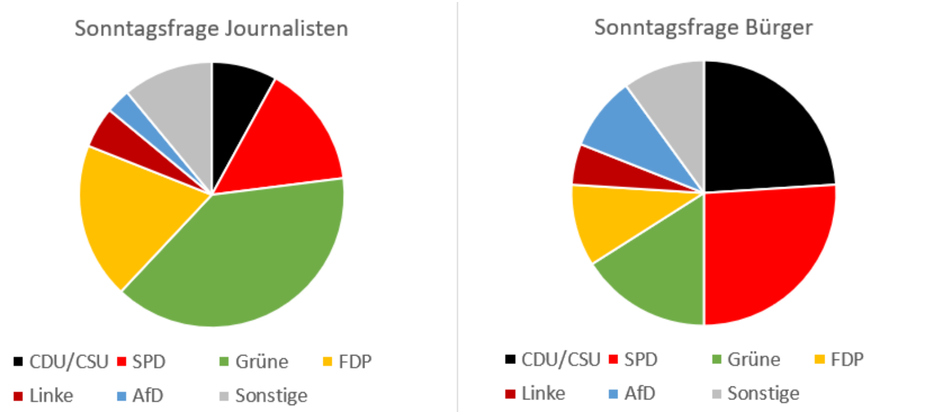



(b) education




*Notes:* Figure B9 illustrates the age distribution (left) and the education levels (right) of the journalist sample.

Figure B10: Voting intentions




*Notes:* Figure B10 summarizes answers to the question of which party one would vote for if there was an election on the following Sunday for the journalist sample (left) and a representative sample for the German population (right). The data of the representative sample comes from the market research institute *forsa*.

Table B6: Balance Table Journalists

| | flat pay (1) | pay-per-click (2) | t-test (3) |
|---|---|---|---|
| age | 46.830 | 46.179 | 0.651 |
| | (1.226) | (1.317) | |
| seniority | 15.217 | 14.284 | 0.933 |
| | (0.914) | (0.964) | |
| risk | 7.613 | 7.968 | -0.355 |
| | (0.190) | (0.212) | |
| patience | 8.151 | 8.179 | -0.028 |
| | (0.169) | (0.205) | |
| altruism | 8.575 | 8.979 | -0.403 |
| | (0.186) | (0.189) | |
| trust | 5.226 | 4.916 | 0.311 |
| | (0.216) | (0.247) | |
| narcissism | 2.679 | 2.937 | -0.258 |
| | (0.123) | (0.152) | |
| *education* | | | |
| secondary school | 9.4% | 12.6% | -3.2% |
| high school | 23.6% | 16.8% | 6.7% |
| university | 64.2% | 67.4% | -3.2% |
| other | 2.8% | 3.2% | -0.3% |
| *political preference* | | | |
| SPD | 15.1% | 12.6% | 2.5% |
| CDU/CSU | 6.6% | 9.5% | -2.9% |
| Die Grünen | 36.8% | 36.8% | 0% |
| FDP | 17.0% | 20.0% | -3.0% |
| AfD | 1.9% | 3.2% | -1.3% |
| Die Linke | 3.8% | 5.3% | -1.5% |
| other | 16% | 6.3% | 9.7%** |
| wouldn't vote | 2.8% | 6.3% | -3.5% |
| phone use | 31.1% | 32.6% | -1.5% |
| online medium | 33.0% | 29.3% | 3.7% |
| Observations | 106 | 95 | |

*Notes:* seniority: aged worked as a journalist; education: highest educational diploma; political preference: percentage of people who would vote for a certain party if there were national elections on the next Sunday; risk, patience and altruism: survey measures of these preferences validated by Falk et al. (2018) on a 11-point Likert scale; narcissism: self-evaluated narcissism on a 7-point Likert scale; phone use: percentage of participants answering the survey on their phone; online medium: percentage of participants working only for an online medium.
The value displayed for t-tests in column 3 are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

## B.3. Main Outcomes

Table B7: OLS Estimates - ATE on Headline Choice

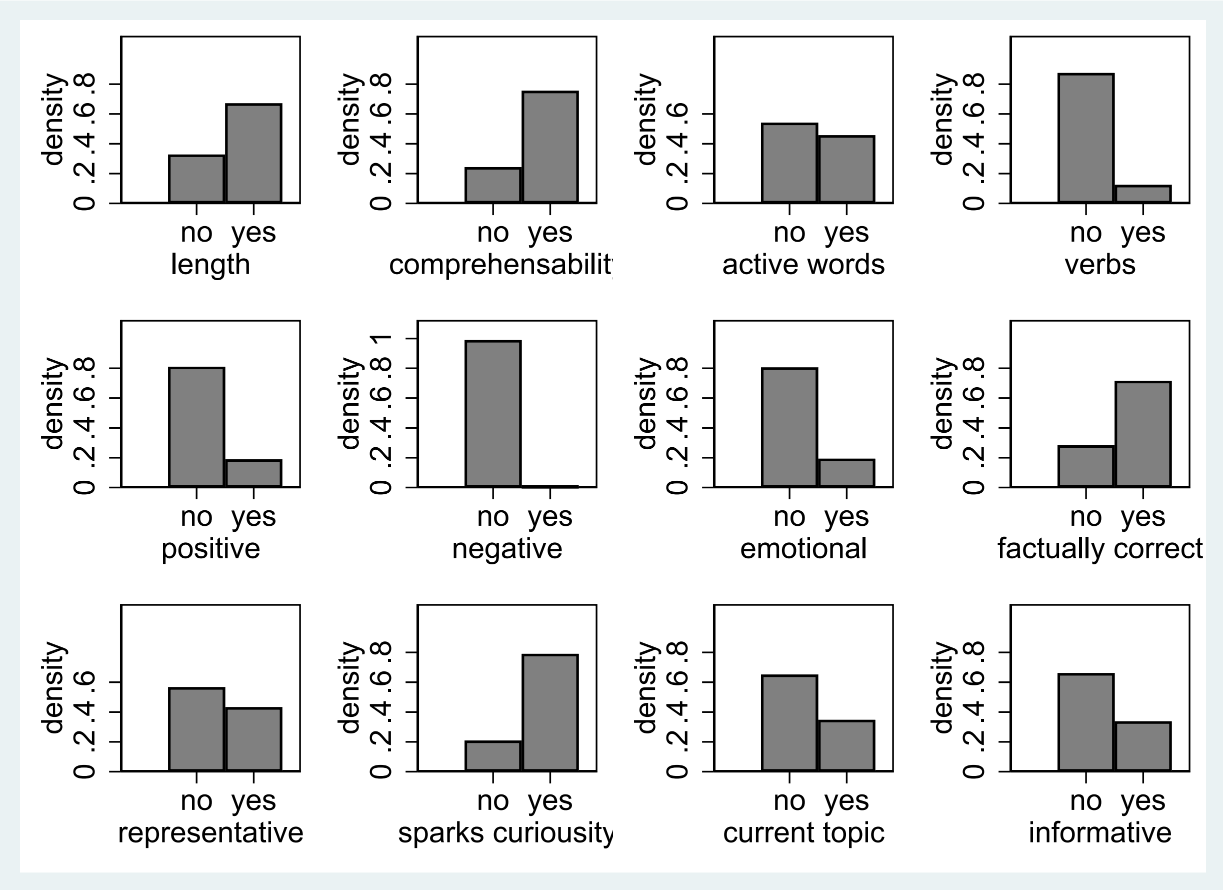
| | Sentiment | | Emotions | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| sentiment | -0.0064548 | 0.0058872 | | |
| | (0.0971431) | (0.1000453) | | |
| emotionality | | | 0.2048659*** | 0.2082093*** |
| | | | (0.0694661) | (0.070733) |
| seniority | | 0.0066359 | | 0.0055346 |
| | | (0.0063493) | | (0.0047168) |
| age | | 0.0072019 | | 0.0036173 |
| | | (0.0052546) | | (0.0039627) |
| vote: SPD | | 0.0265586 | | 0.2931497 |
| | | (0.2956812) | | (0.2892631) |
| vote: CDU/CSU | | 0.1154226 | | 0.2760218 |
| | | (0.2991308) | | (0.3052485) |
| vote: Die Grünen | | 0.076108 | | 0.444624 |
| | | (0.2815415) | | (0.287009) |
| vote: FDP | | 0.0645764 | | 0.2155939 |
| | | (0.276671) | | (0.2828545) |
| vote: Die Linke | | 0.1567016 | | 0.1004916 |
| | | (0.3168345) | | (0.3139044) |
| vote: other | | -0.1524632 | | 0.2851998 |
| | | (0.312055) | | (0.2988047) |
| vote: wouldn't vote | | -0.160409 | | 0.4430629 |
| | | (0.3906121) | | (0.3253319) |
| role: producer | | -0.0316488 | | 0.0754752 |
| | | (0.1369495) | | (0.1224299) |
| role: graphs | | 0.1340798 | | 0.154554 |
| | | (0.1634986) | | (0.1400021) |
| role: other | | -0.2236172 | | 0.0432215 |
| | | (0.2368406) | | (0.1883472) |
| education: university | | -0.0340787 | | 0.1098455 |
| | | (0.1208012) | | (0.0854872) |
| education: high school | | 0.2075383 | | 0.3093532** |
| | | (0.2036895) | | (0.1417009) |
| education: other | | 0.1298381 | | 0.0022261 |
| | | (0.2124546) | | (0.180188) |
| phone use | | 0.069182 | | 0.0551961 |
| | | (0.1097543) | | (0.0778126) |
| risk preference | | 0.0257794 | | 0.0297685 |
| | | (0.0287157) | | (0.0198317) |
| patience | | 0.025475 | | 0.0200179 |
| | | (0.0294635) | | (0.01995) |
| altruism | | 0.0003739 | | -0.0131103 |
| | | (0.0294723) | | (0.0231197) |
| trust | | 0.0456064* | | 0.014938 |
| | | (0.0231822) | | (0.0165412 |
| narcissism | | -0.0629251 | | -0.0221183 |
| | | (0.038211) | | (0.0259116) |
| Constant | 0.2169811*** | -0.7394713 | 0.4056604*** | -0.6645239 |
| | (0.0584536) | (0.5248686) | (0.0479311) | (0.4230708) |
| $R^2$ | 0.0000 | 0.1280 | 0.0418 | 0.1555 |
| Observations | 201 | 201 | 201 | 201 |

*Notes:* Table B7 reports OLS estimates with robust standard errors in parentheses. The flat rate group is the reference group. For vote the omitted category is voting for the AfD. For role the omitted category is editing journalistic pieces. For education the omitted category is A-levels (Abitur). *** p<0.01, ** p<0.05, * p<0.1

## B.4. Secondary Outcomes

### General factors for composing a headline

Figure B11: General characteristics for headlines




*Notes:* Figure B11 summarizes answers to the question of which factors journalist regard most important for composing a headline *in general* (not in the context of this study or certain incentives). The distributions are across both treatment groups. There is no statistically significant difference between the two groups for any of the factors on the 5 or 1 percent level. On the 10 percent level the XX factor is slightly more important for the X group (p=X).

## B.5. Additional Analyses on Mechanisms

### B.5.1. Descriptive Evidence on Headline Competition in Offline Markets

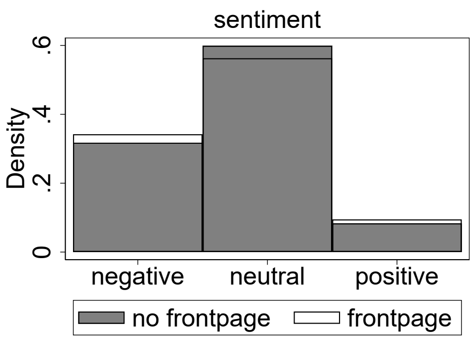
Does higher competition drive incentives to select emotional headlines? This section serves as a consistency check: If being published on the frontpage is correlated with higher emotionality of a headline, this makes the the hypothesis that the degree of competition is related to selecting emotional headlines more plausible.

The offline data of the main dataset does not come with information of the specific page a article was published on. This information is available in the offline part of the robustness datasets, therefore they are being used here. This means I can compare frontpage to non-frontpage headlines for *The New York Times*, *Der Spiegel* and *BILD*. The empirical strategy is identical to the one described in equations 1 to 5, but instead of an online dummy a dummy for a headline being on the frontpage or not is being used. Additionally - just as in the robustness section - as these datasets do not contain the article content, I cannot control for content tonality. Note that any evidence presented here is again correlational, and not necessarily causal.

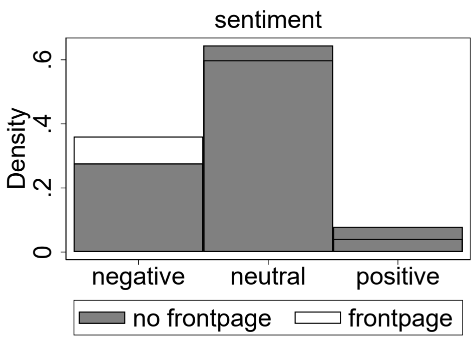
The comparison shows that frontpage headlines are indeed more emotional and more negative than headlines of stories in other parts of the papers. For *The New York Times*, frontpage headlines are on average 0.26 standard deviations more emotional and 0.18 standard deviations more negative than headlines on other pages. For *Der Spiegel*, frontpage headlines are on average 0.09 standard deviations more emotional and 0.18 standard deviations more negative than non-frontpage headlines. For *BILD*, frontpage headlines are on average 0.07 standard deviations more emotional and 0.02 standard deviations more negative than headlines in other parts of the newspaper. Figure B12 illustrates the distribution of the headline classifications for all three datasets. Regression results are available in Tables B8 - B13.

Figure B12: Sentiment of Frontpage- and Non-Frontpage Headlines

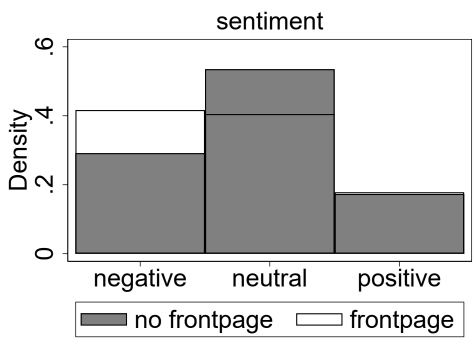
(a) BILD




(b) Der Spiegel




(c) The New York Times




*Notes:* Figures B12(a) - B12(c) illustrate the distribution of the sentiment variable for the offline data in the different datasets. Headlines that were not published on the frontpage are shaded in gray.

Table B8: OLS Estimates - Emotionality of BILD Headlines

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| online | 0.0725*** | 0.0227*** | 0.0706*** | 0.0154* |
| | (0.0084) | (0.0086) | (0.0084) | (0.0087) |
| article length | | | | -0.0368** |
| | | | | (0.0133) |
| topic FE | no | yes | no | yes |
| time FE | no | no | yes | yes |
| Constant | -0.0972*** | -0.0472*** | -0.1418*** | -0.1264*** |
| | (0.0034) | (0.0039) | (0.0305) | (0.0305) |
| $R^2$ | 0.0008 | 0.0077 | 0.0038 | 0.0124 |
| Observations | 97,576 | 97,576 | 97,576 | 97,576 |

*Notes:* Table B8 reports OLS estimates in standard deviations with robust standard errors in parentheses. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.2. The non-front-page headlines are always the reference group. Control variables are article length in 1000 words and fixed effects as dummies for the different topics and days. *** p<0.01, ** p<0.05, * p<0.1

Table B9: OLS Estimates - Sentiment of BILD Headlines

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| online | -0.0211** | 0.0270*** | -0.0207** | 0.0339*** |
| | (0.0083) | (0.0086) | (0.0083) | (0.0085) |
| article length | | | | -0.0194 |
| | | | | (0.0122) |
| topic FE | no | yes | no | yes |
| time FE | no | no | yes | yes |
| Constant | 0.0420*** | -0.0064* | 0.1094*** | 0.1043*** |
| | (0.0033) | (0.0038) | (0.0296) | (0.0297) |
| $R^2$ | 0.0001 | 0.0069 | 0.0020 | 0.0099 |
| Observations | 97,576 | 97,576 | 97,576 | 97,576 |

*Notes:* Table B9 reports OLS estimates with robust standard errors in parentheses. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.2. The non-front-page headlines are always the reference group. Control variables are article length in 1000 words and fixed effects as dummies for the different topics and days. *** p<0.01, ** p<0.05, * p<0.1

Table B10: OLS Estimates - Emotionality of Spiegel Headlines

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| online | 0.0929 | 0.0628 | 0.0861 | 0.0788 |
| | (0.0827) | (0.0846) | (0.0820) | (0.0892) |
| article length | | | | -0.0106 |
| | | | | (0.0131) |
| topic FE | no | yes | no | yes |
| time FE | no | no | yes | yes |
| Constant | -0.4178*** | -0.5526*** | -0.4297*** | -0.5401*** |
| | (0.0140) | (0.0359) | (0.0433) | (0.0537) |
| $R^2$ | 0.0003 | 0.0347 | 0.0030 | 0.0372 |
| Observations | 4,896 | 4,896 | 4,896 | 4,896 |

*Notes:* Table B10 reports OLS estimates in standard deviations with robust standard errors in parentheses. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.2. The non-front-page headlines are always the reference group. Control variables are article length in 1000 words and fixed effects as dummies for the different topics and days. *** p<0.01, ** p<0.05, * p<0.1

Table B11: OLS Estimates - Sentiment of Spiegel Headlines

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| online | -0.1778*** | -0.1169* | -0.1663** | -0.0471 |
| | (0.0668) | (0.0682) | (0.0664) | (0.0724) |
| article length | | | | -0.0242** |
| | | | | (0.0110) |
| topic FE | no | yes | no | yes |
| time FE | no | no | yes | yes |
| Constant | 0.1535*** | 0.1540*** | 0.1943*** | 0.1965*** |
| | (0.0118) | (0.0286) | (0.0371) | (0.0449) |
| $R^2$ | 0.0014 | 0.0241 | 0.0035 | 0.0273 |
| Observations | 4,896 | 4,896 | 4,896 | 4,896 |

*Notes:* Table B11 reports OLS estimates in standard deviations with robust standard errors in parentheses. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.2. The non-front-page headlines are always the reference group. Control variables are article length in 1000 words and fixed effects as dummies for the different topics and days. *** p<0.01, ** p<0.05, * p<0.1

Table B12: OLS Estimates - Emotionality of NYT Headlines

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| online | 0.2602*** | 0.1351*** | 0.2603*** | 0.1201*** |
| | (0.0165) | (0.0163) | (0.0165) | (0.0171) |
| article length | | | | 0.0231*** |
| | | | | (0.0082) |
| topic FE | no | yes | no | yes |
| time FE | no | no | yes | yes |
| Constant | -0.0542*** | -0.7082*** | -0.0292 | -0.7083*** |
| | (0.0062) | (0.0906) | (0.0185) | (0.0913) |
| $R^2$ | 0.0082 | 0.1233 | 0.0086 | 0.1242 |
| Observations | 29,216 | 29,216 | 29,216 | 29,216 |

*Notes:* Table B12 reports OLS estimates in standard deviations with robust standard errors in parentheses. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.2. The non-front-page headlines are always the reference group. Control variables are article length in 1000 words and fixed effects as dummies for the different topics and days. *** p<0.01, ** p<0.05, * p<0.1

Table B13: OLS Estimates - Sentiment of NYT Headlines

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| online | -0.1755*** | -0.0993*** | -0.1756*** | -0.0668*** |
| | (0.0178) | (0.0182) | (0.0178) | (0.0189) |
| article length | | | | -0.0512*** |
| | | | | (0.0073) |
| topic FE | no | yes | no | yes |
| time FE | no | no | yes | yes |
| Constant | 0.0613*** | 0.4368*** | 0.0489** | 0.0489*** |
| | (0.0062) | (0.0663) | (0.0186) | (0.0186) |
| $R^2$ | 0.0037 | 0.0749 | 0.0041 | 0.0765 |
| Observations | 29,216 | 29,216 | 29,216 | 29,216 |

*Notes:* Table B13 reports OLS estimates in standard deviations with robust standard errors in parentheses. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.2. The non-front-page headlines are always the reference group. Control variables are article length in 1000 words and fixed effects as dummies for the different topics and days. *** p<0.01, ** p<0.05, * p<0.1

### B.5.2. Anecdotal Evidence on Headline Competition in Online Markets

A plausible reason for increased incentives to capture readers' attention in online markets is higher competition on the headline level, which in turn might translate into more emotional headlines. To explore whether the degree of competition can really be related to the tonality of headlines I copy the idea behind the identification strategy of Meyer et al. (2022) and leverage the exclusion of a group of news outlets from popular news aggregators due to a legal dispute in Germany.

**Context** The exogenous variation in competition that I use is an exclusion of one of the newspapers in my sample, namely *Die Welt*, from several news aggregators due to a legal dispute in 2014. News aggregators are websites that combine news content from several sources, typically summarize the articles' content in one paragraph and provide the link to the original. Especially in the early days of these aggregators, traditional news companies oftentimes regarded them as a threat. They were afraid that many readers would be satisfied with the small summaries and thus use the aggregators as a substitute to consumption of online news from traditional providers. In this context, the German government passed the "ancillary copyright for press publishers". This bill allows traditional media companies to claim royalty fees if their content is used by other companies. However, short excerpts of text were excepted from the regulation, which led to uncertainty about if the new legislation applied to the paragraphs that news aggregators provided. The German copyright collecting society *VGM* published a pricing schedule for the reuse of its members' original content and threatened to file lawsuits against news aggregators that refused to pay[40]. As a response, a group of popular news aggregators (for example *gmx.de*, *web.de*, and *t-online.de*) removed all articles from of *VGM* members from their platforms in August 2014. They continued to show content from non-members. Thus, competition on the headline level decreased for *VGM* members, but not for non-members. I use this removal of *Die Welt* to explore how changes in competition can translate into changes of tonality.

**Empirical Strategy** Note that I only have data on three news outlets for the relevant time frame[41], namely *Der Spiegel*, *Die Welt* and *Die Zeit*. Out of the three available outlets, *Die Welt* was a *VGM* member, while the others were not. With that data I conduct two comparisons. First, I regard the development of the online versions *Die Welt* in contrast to the online versions of the other newspapers before and after the removal. Second, I compare the tonality of the online headlines to the offline headlines of *Die Welt* before and after the removal.

The copyright bill was passed in March 2013 and the *VGM* members were removed from the aggregators in August 2014. As suggested in the empirical strategy of Meyer et al. (2022), I drop the time frame between these two dates to exclude potential confounders that may exist because of the ongoing debate about the legislation. Therefore, the regarded time frames are always a comparison of before March 2013 to after August 2014. In line with Meyer et al. (2022) I limit the time frame to 18 months before and after the removal, but also assess different time frames as a robustness checks later on. I graphically analyze the results and compute a back-of-the-envelope difference in differences estimator to get an idea about the effect size. Because of the low number of outlets in my sample (which hampers clustering standard errors on the outlet level), I don't run regressions to obtain this estimate.

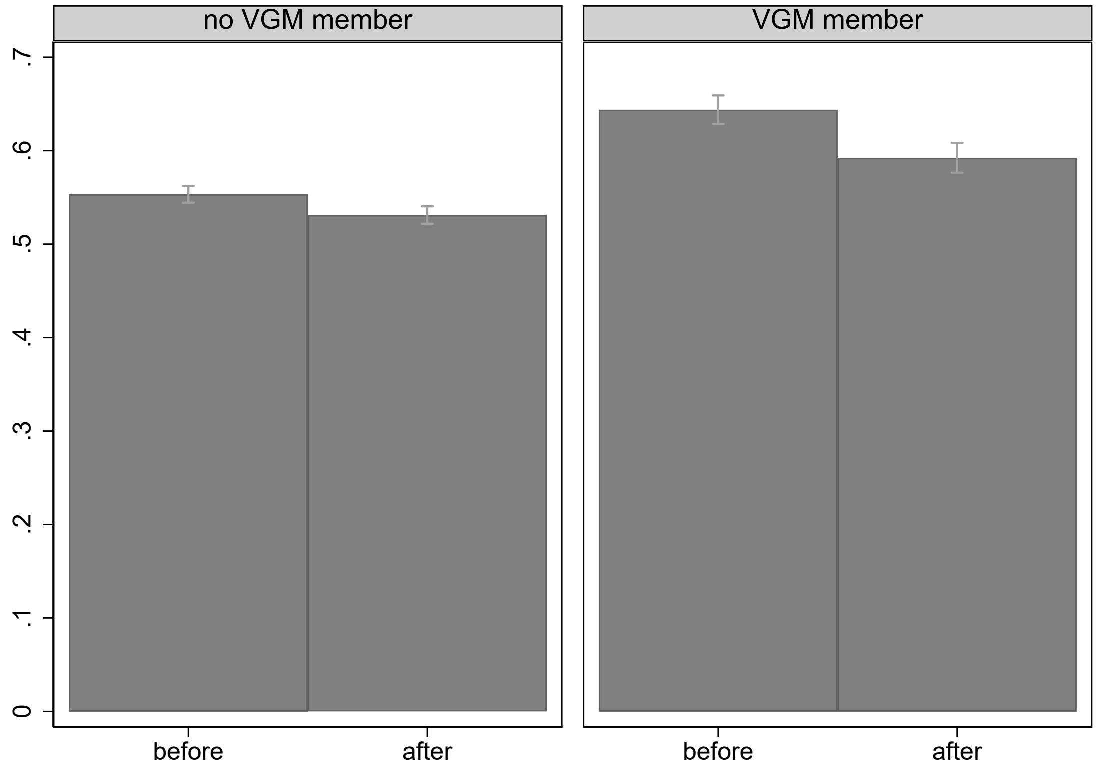
**Main Finding** The comparison reveals that the removal from the aggregators seems to have slightly reduced the share of emotional headlines at *Welt Online*. Figure B13 illustrates that comparison. Back-of-the-envelope estimations of the difference in differences suggest that the effect size is around 0.08 or 0.13 standard deviations (depending on which control group is used). I don't detect differences in terms of sentiment, as depicted in Figure B16.

[40]See this press article for a summary of the discussion.

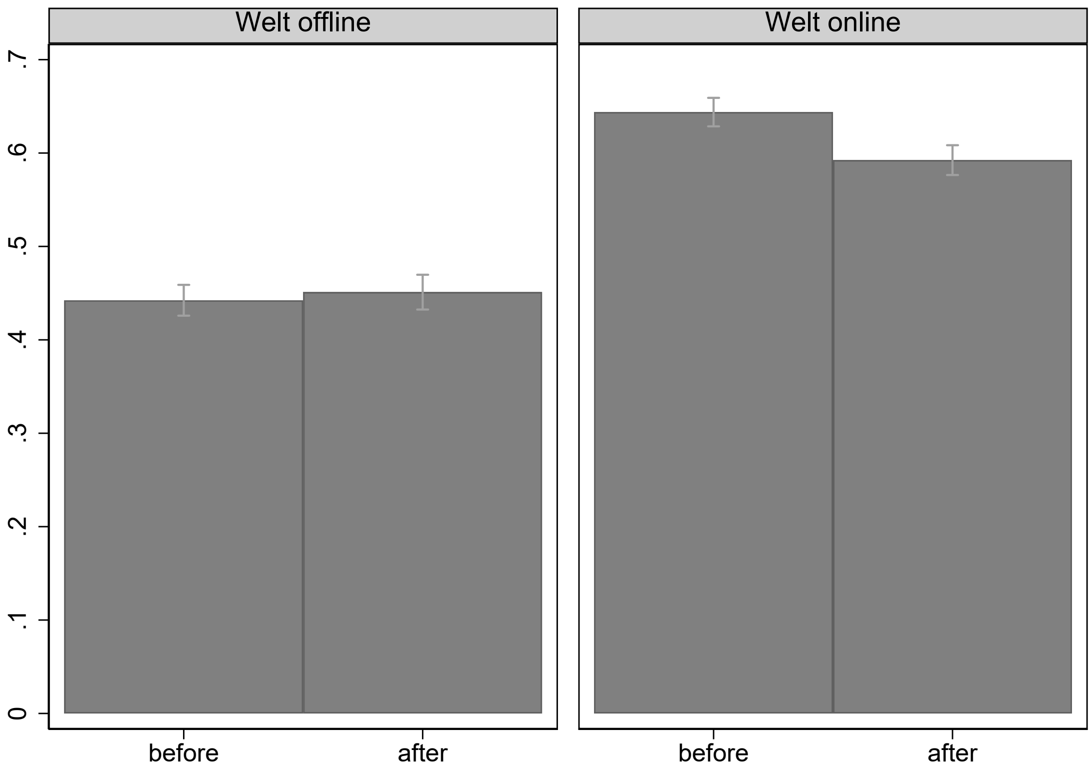
[41]A detailed description of the availability of the data at different points in time is denoted in Table 1.

Figure B13: Comparing the emotionality of headlines in the removed outlet to others

(a) Emotionality of the online headlines of the removed and non-removed outlets



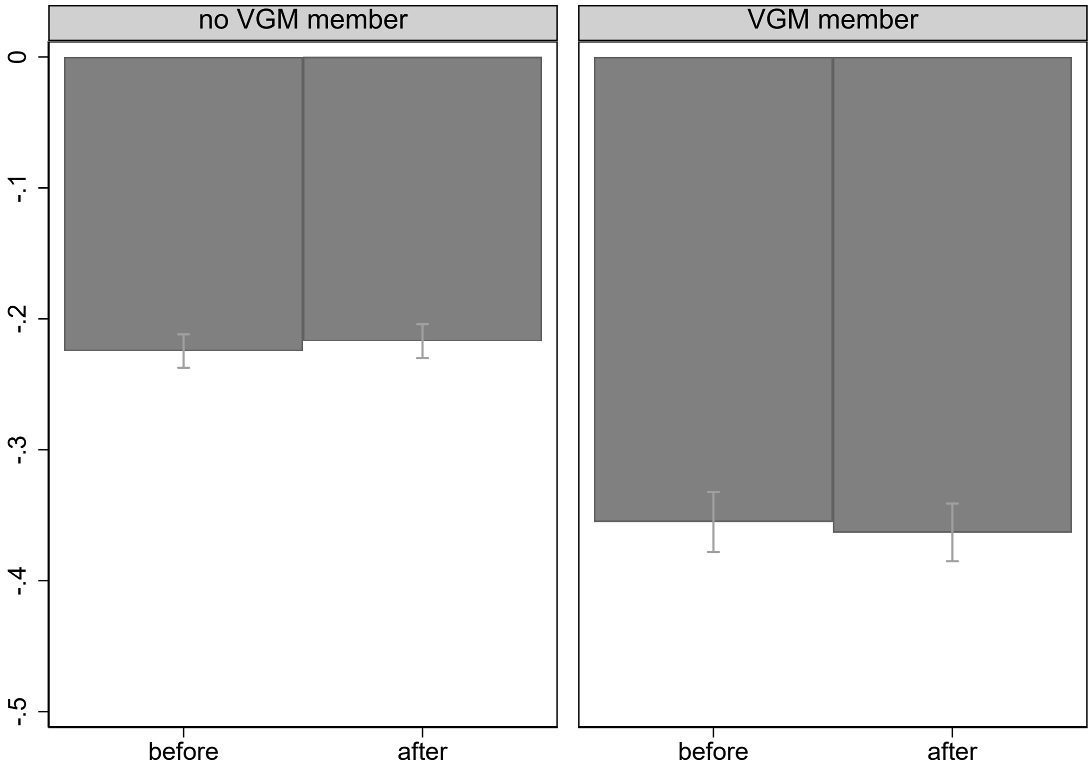

(b) Emotionality of the online and offline headlines of the removed outlet



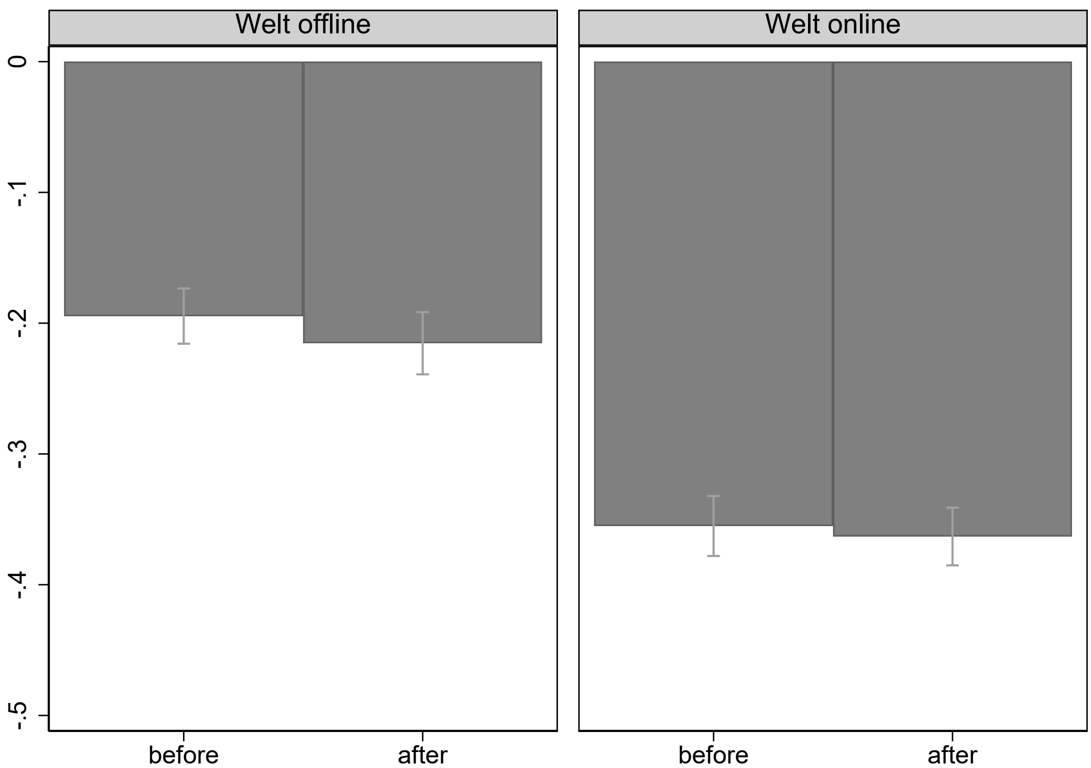

*Notes:* Figure B13(a) depicts the average emotionality of headlines at VGM members and non-members before and after the removal. Figure B13(b) illustrates the average emotionality of headlines at the online and offline versions of *Die Welt* before and after the removal. The time frame used for computing the averages is always 18 months before March 2013 and 18 months after August 2014.

Figure B14: Comparing the sentiment of headlines in the removed outlet to others

(a) Sentiment of the online headlines of the removed and non-removed outlets




(b) Sentiment of the online and offline headlines of the removed outlet




*Notes:* Figure B16(a) depicts the average sentiment of headlines at VGM members and non-members before and after the removal. Figure B16(b) illustrates the average sentiment of headlines at the online and offline versions of *Die Welt* before and after the removal. The time frame used for computing the averages is always 18 months before March 2013 and 18 months after August 2014.

**Robustness: Alternative depended variables** I assess the robustness of the finding by using the classifications of the other sentiment classifiers as described in Appendix A.1 as alternative outcomes. The finding that the removal reduces emotionality reproduces qualitatively with all other classification methodologies. The size of the estimated effect ranges from 0.05 to 0.17 standard deviations. Illustrations are available in Figures B15 to B18.

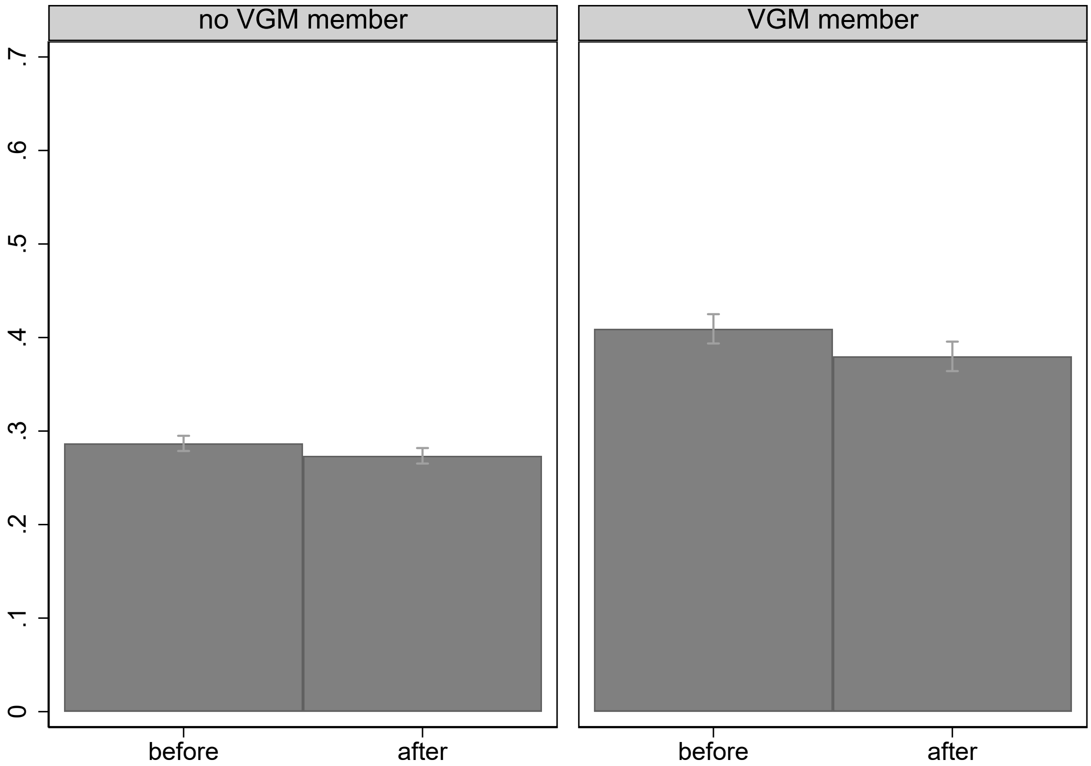
**Robustness: Alternative time windows** A concern could be that the result depends on the specific time frame used. To explore this, I repeat the comparison, but instead of a 18 months I change the time frame to 24 or 12 months before and after the removal. Also, I repeat the main analysis without removing the period from March 2013 to August 2014 from the dataset. These are the same robustness time frames that Meyer et al. (2022) test.

The result is robust to both an expansion and limitation of the regarded time frames. If the time frame is reduced to one year before and after the removal the estimate for the difference in differences is $-0.06$ standard deviations. When it is expanded to 2 years the resulting estimate is $-0.08$ standard deviations. Including the time frame between March 2013 to August 2014 in the analysis yields an estimate of $-0.06$ standard deviations. Illustrations of the differences with the modified time-frames are available in Figures B19 to B21.

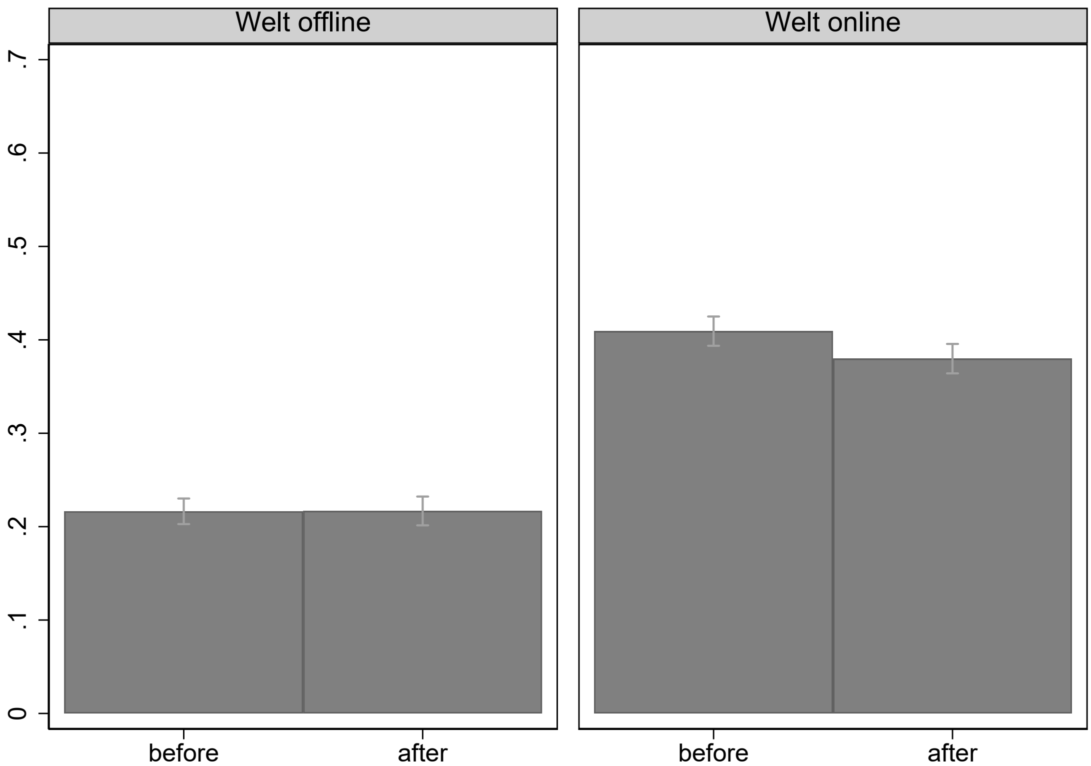
**Limitations** As previously mentioned, the number of outlets used in this analysis is limited to three. Therefore, I recommend interpreting the results as first, suggestive evidence for the described mechanism. It seems to be a promising avenue for future research to further explore this relationship with more diverse samples.

Figure B15: Robustness: LM dictionary as alternative classifier

(a) Emotionality of the online headlines of the removed and non-removed outlets



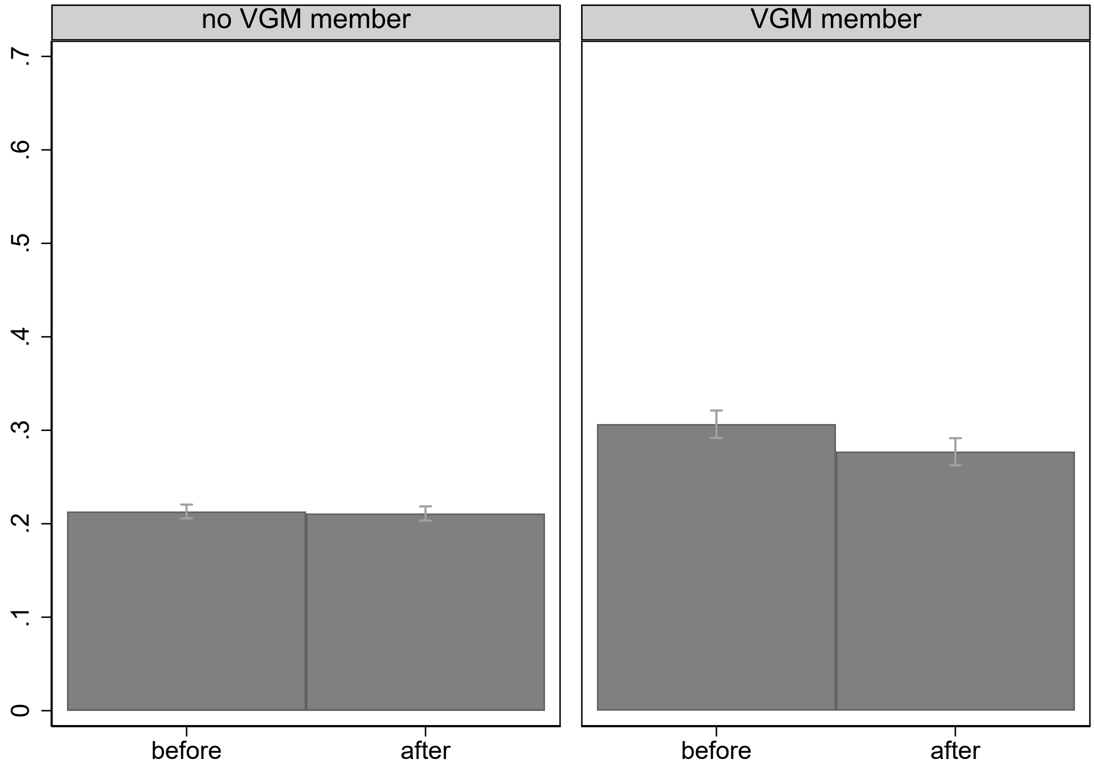

(b) Emotionality of the online and offline headlines of the removed outlet



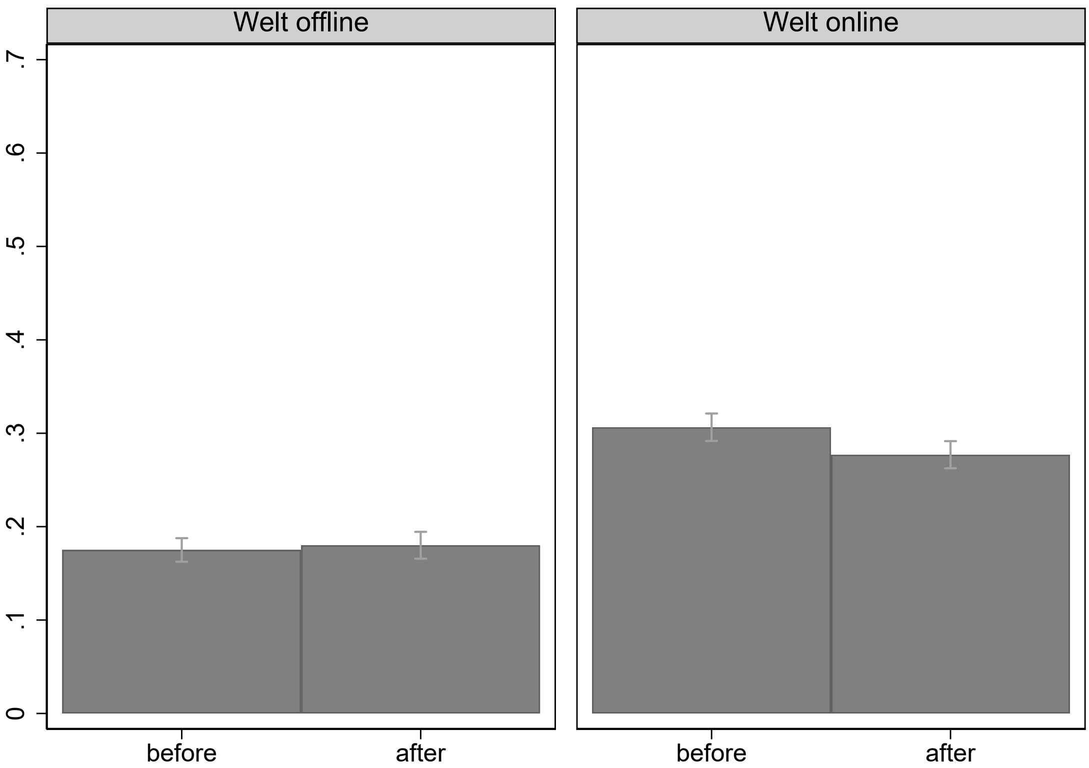

*Notes:* Figure B15(a) depicts the average sentiment of headlines at VGM members and non-members before and after the removal. Figure B15(b) illustrates the average sentiment of headlines at the online and offline versions of *Die Welt* before and after the removal. The time frame used for computing the averages is always 18 months before March 2013 and 18 months after August 2014.

Figure B16: Robustness: SentiWS dictionary as alternative classifier

(a) Emotionality of the online headlines of the removed and non-removed outlets



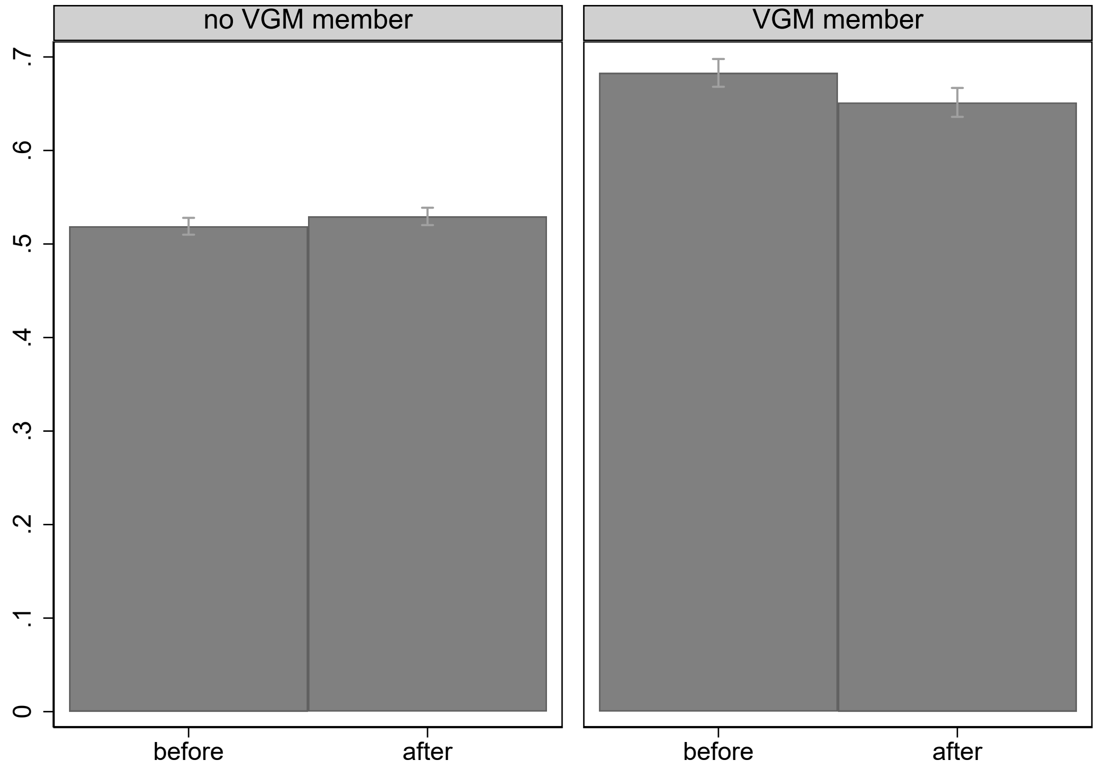

(b) Emotionality of the online and offline headlines of the removed outlet



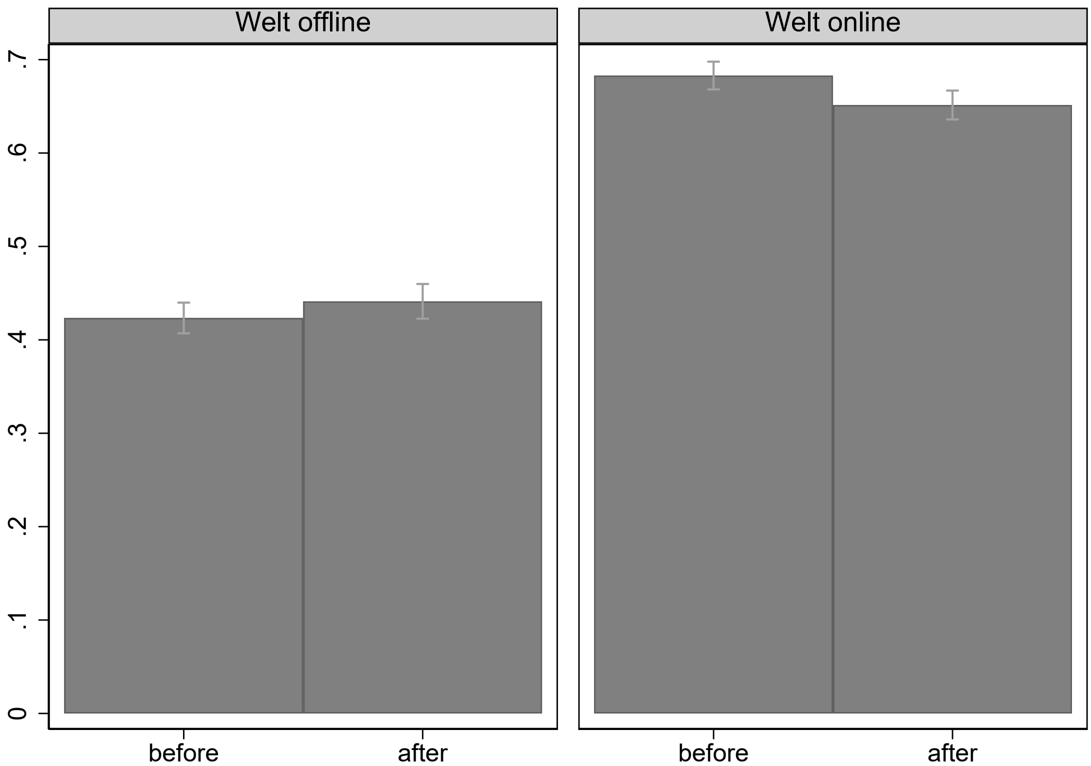

*Notes:* Figure B16(a) depicts the average sentiment of headlines at VGM members and non-members before and after the removal. Figure B16(b) illustrates the average sentiment of headlines at the online and offline versions of *Die Welt* before and after the removal. The time frame used for computing the averages is always 18 months before March 2013 and 18 months after August 2014.

Figure B17: Robustness: VADER dictionary as alternative classifier

(a) Emotionality of the online headlines of the removed and non-removed outlets



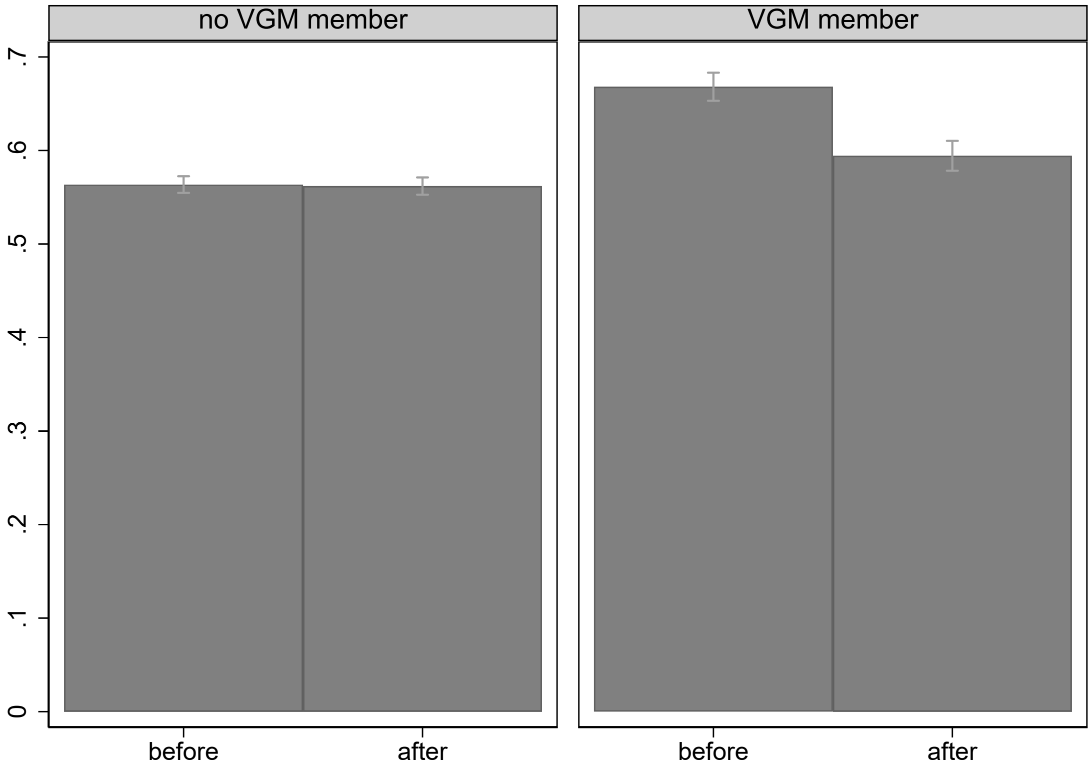

(b) Emotionality of the online and offline headlines of the removed outlet



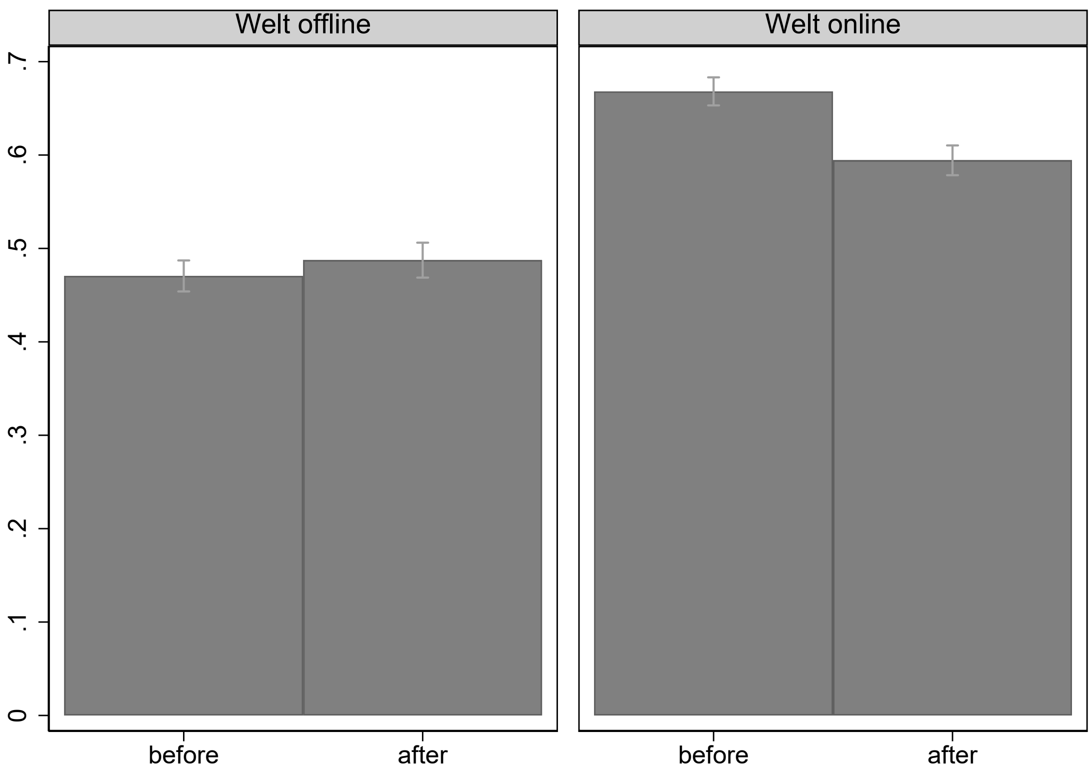

*Notes:* Figure B17(a) depicts the average sentiment of headlines at VGM members and non-members before and after the removal. Figure B17(b) illustrates the average sentiment of headlines at the online and offline versions of *Die Welt* before and after the removal. The time frame used for computing the averages is always 18 months before March 2013 and 18 months after August 2014.

Figure B18: Robustness: Non-tuned roBERTa as alternative classifier

(a) Emotionality of the online headlines of the removed and non-removed outlets



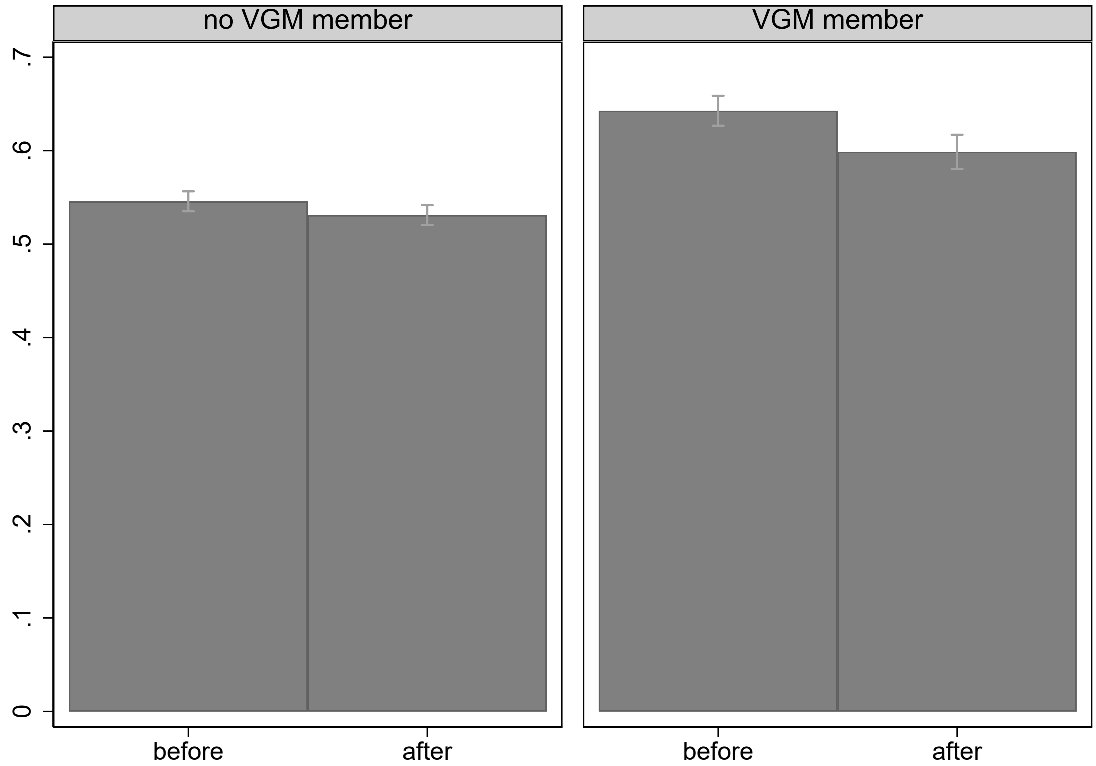

(b) Emotionality of the online and offline headlines of the removed outlet



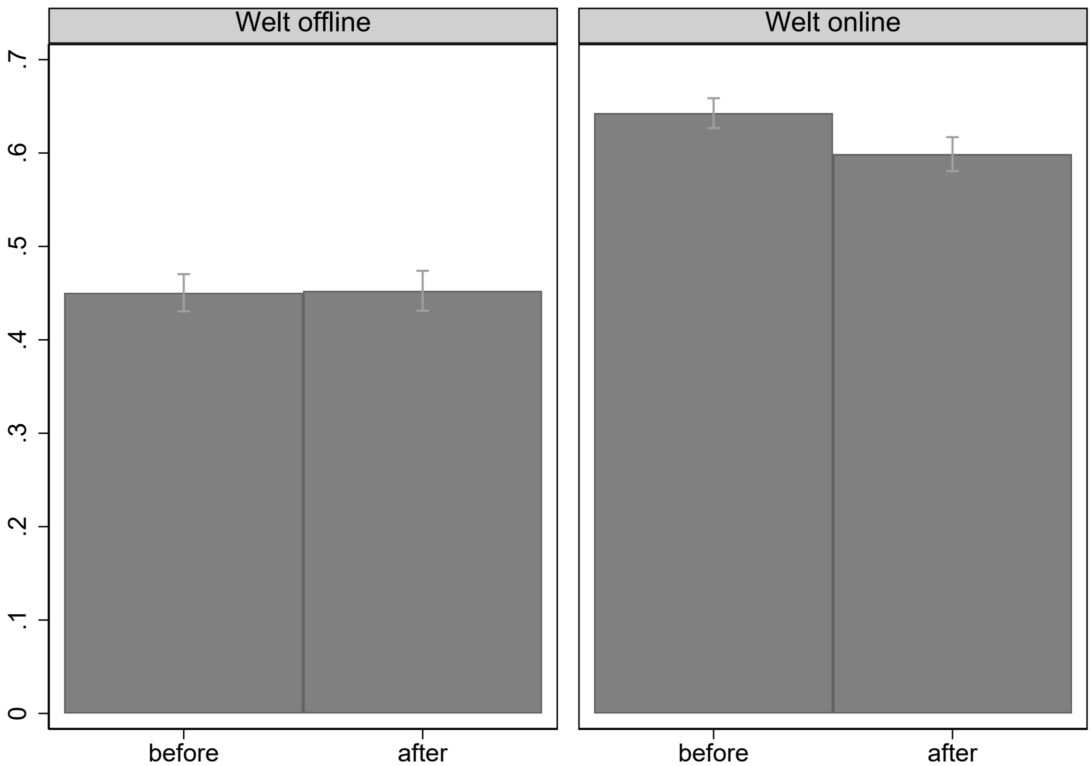

*Notes:* Figure B18(a) depicts the average sentiment of headlines at VGM members and non-members before and after the removal. Figure B18(b) illustrates the average sentiment of headlines at the online and offline versions of *Die Welt* before and after the removal. The time frame used for computing the averages is always 18 months before March 2013 and 18 months after August 2014.

Figure B19: Robustness: Time window of 12 months

(a) Emotionality of the online headlines of the removed and non-removed outlets



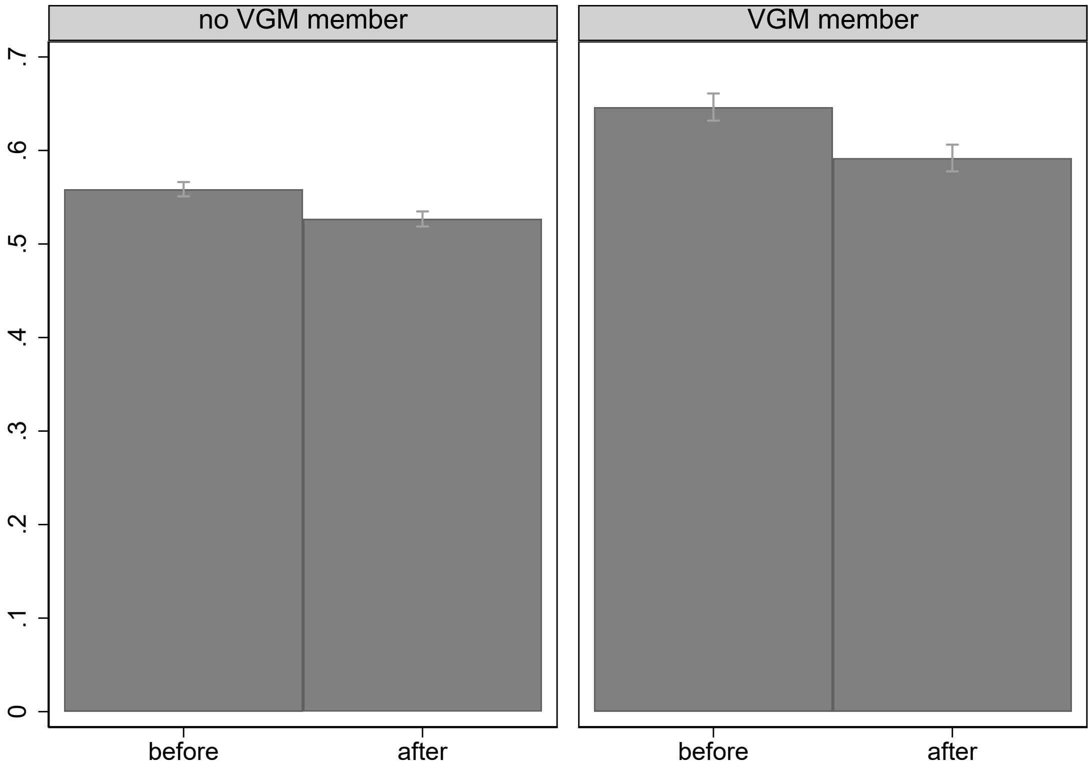

(b) Emotionality of the online and offline headlines of the removed outlet



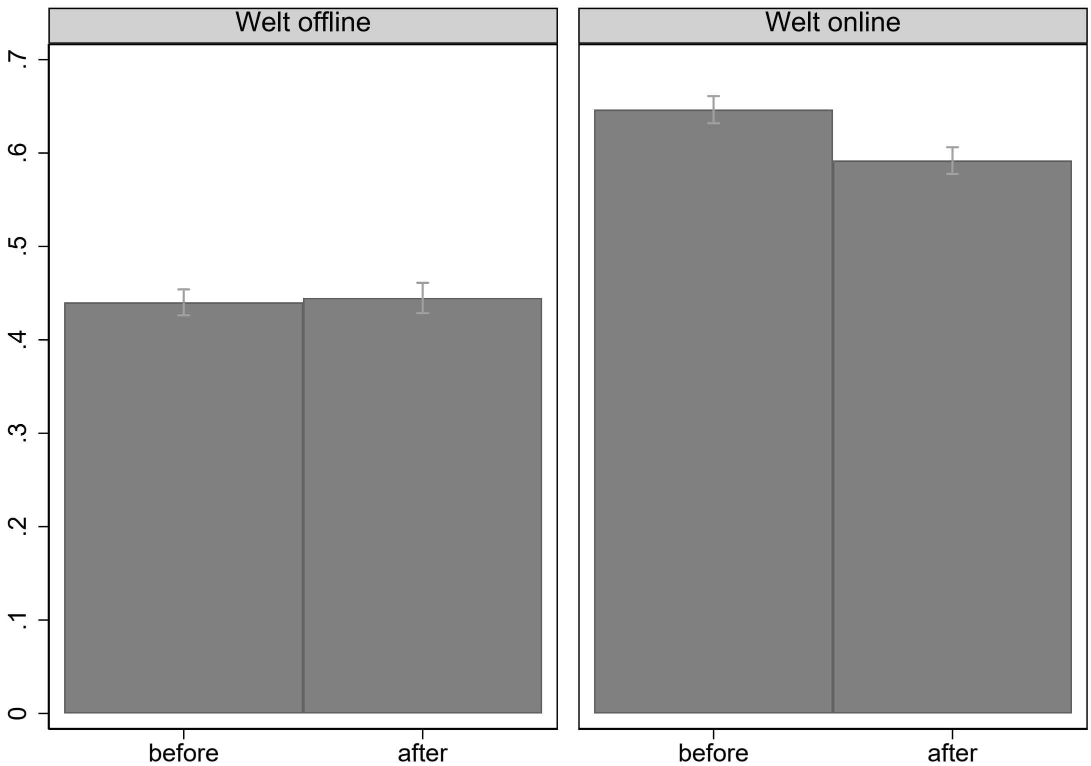

*Notes:* Figure B19(a) depicts the average sentiment of headlines at VGM members and non-members before and after the removal. Figure B19(b) illustrates the average sentiment of headlines at the online and offline versions of *Die Welt* before and after the removal. The time frame used for computing the averages is always 12 months before March 2013 and 12 months after August 2014.

Figure B20: Robustness: Time window of 24 months

(a) Emotionality of the online headlines of the removed and non-removed outlets



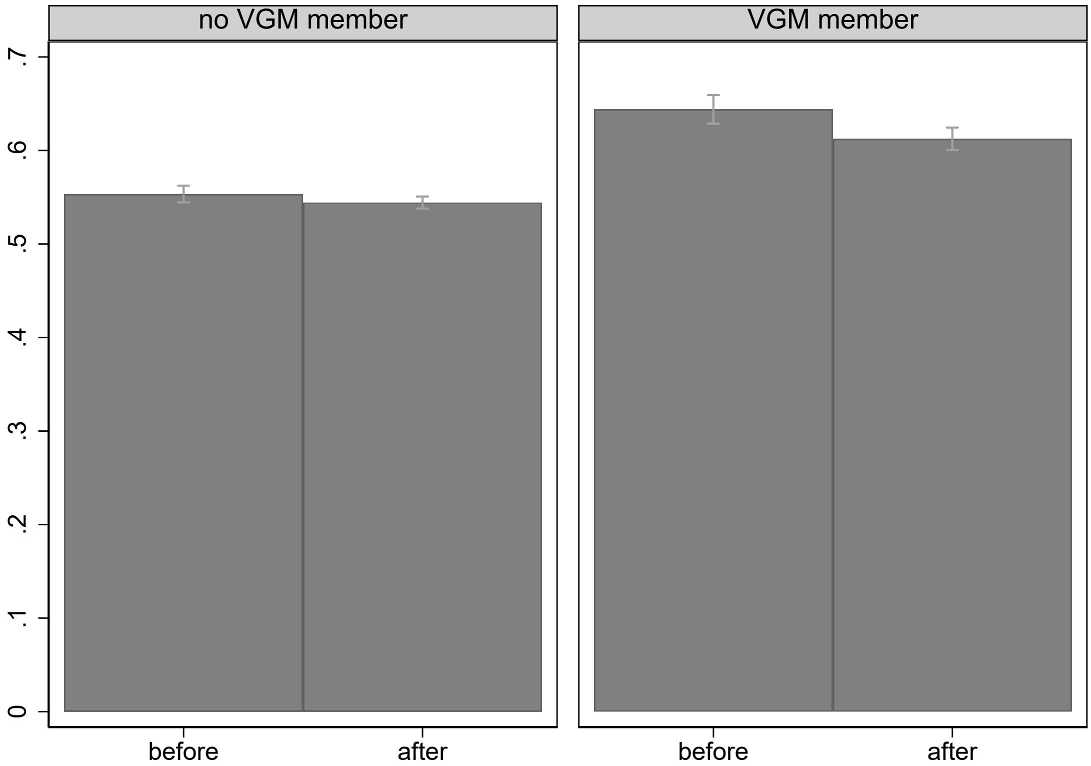

(b) Emotionality of the online and offline headlines of the removed outlet



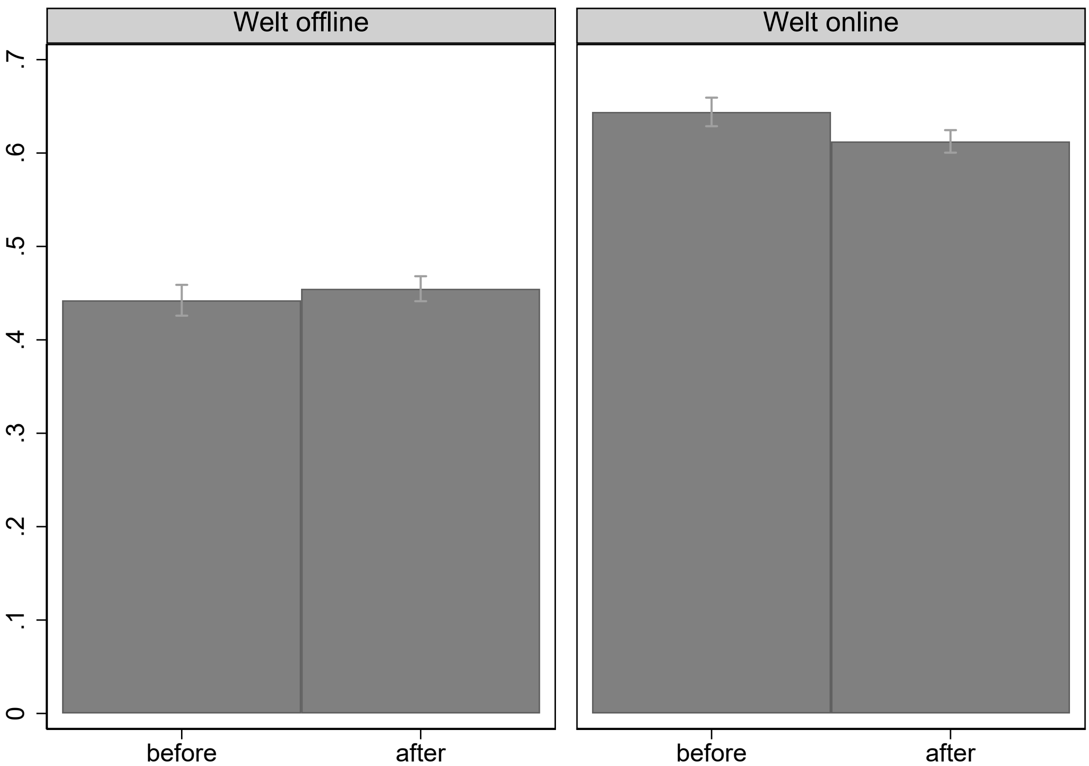

*Notes:* Figure B20(a) depicts the average sentiment of headlines at VGM members and non-members before and after the removal. Figure B20(b) illustrates the average sentiment of headlines at the online and offline versions of *Die Welt* before and after the removal. The time frame used for computing the averages is always 24 months before March 2013 and 24 months after August 2014.

Figure B21: Robustness: Time window without gap

(a) Emotionality of the online headlines of the removed and non-removed outlets




(b) Emotionality of the online and offline headlines of the removed outlet




*Notes:* Figure B21(a) depicts the average sentiment of headlines at VGM members and non-members before and after the removal. Figure B21(b) illustrates the average sentiment of headlines at the online and offline versions of *Die Welt* before and after the removal. The time frame used for computing the averages is always 18 months before March 2013 and 18 months after August 2014, without the time frame between those two points in time being dropped.

### B.5.3. Automated Finance News at Welt Online in 2018

With advances in language generation models and increasing automation in many industries automated news - i.e. articles written by computers and not by humans - have become a hot topic in the news industry. They are able to produce a large number of texts in much shorter time than human reporters. However, the type of tasks for which they create value are (at least for what is used in practice up to now) limited to very specific text generation. They are good at reporting about events that occur frequently and always follow similar structures.[42] For example, machines are much faster in summarizing the latest sport results for every small sport club within a country or in writing fine-grained weather forecasts for every region. Publishing these kind of texts only makes sense in the digital sphere, where no space limits exist and each text can reach its' very niche readership.

Automated news are, thanks to their strict structure, oftentimes quite clearly either positive or negative in their framing. For example, for a stock market report, a computer can easily compare today's stock price with yesterdays and write "stock X faced great losses". If it were feasible to have human reporters write about these niche topics (which in practice is far too expensive), they would probably look at a stock price in a broader context and choose a more nuanced headline. It is therefore possible that this type automated reports (partially) drives the difference in the tonality of online and offline headlines described in section 2. This is what this subsection investigates.

**Context** Many outlets have been experimenting with automated news. From the ones in my sample, it is known that *Die Welt* has been piloting this kind of reporting since 2015 and uses it as an integrated part of digital news since 2016.[43] The CEO of Axel Springer[44] confirmed in 2018 that so called "news robots" are being increasingly used and continue to be of growing importance for the newspapers the digital strategy.[45]

In march 2018, *Die Welt* started to use a robot to generate news on finance topics and published them in a quite hidden section of their finance departments subpage[46]. This flooded the ticker with very short news that compared the development of one specific stock or index to another one. The reason that they were published in such a hidden place of the homepage was probably that these articles are targeting a very specific audience, which is more likely to be reached via personalized search engines and social media. I study the introduction of this specific kind of automated news at *Die Welt* to explore which effects automated reporting can have on average headline tonality.

**Empirical Strategy** This event provides a suitable set-up for a difference-in-differences estimation. First of all, the average tonality in the offline versions of the headlines is a good comparison group to the average tonality in the online versions. The reports are talking about the same (more or less emotional) reality and come from the same group of journalists and news outlet. The introduction of automated news however only potentially affects the average tonality of online headlines and leaves offline headlines unaffected.

I use the data on *Die Welt* that was already used as part of the the main descriptive analysis, but limit the analysis to agency content.[47] I limit the regarded time frame to one year before and after the event and compare the average emotionality on the monthly level. I estimate equation

[42] While modern text generation models like GPT2 are also able to produce more creative texts, they are up to now way too prone to make factual errors to be used in the news industry.

[43] See the statements in this interview of an employee of the digital- and innovations unit of the newspaper.

[44] the company owning *Die Welt*

[45] See this report.

[46] WaybackMachine provides an impression of how the ticket looked before this implementation here. An impression of the ticker afterwards is available here.

[47] With my definition of agency content automated news are always classified as agency content. Limiting the analysis to this specific type of news minimizes the risk of picking up some other, unknown change, that might have impacted the tonality journalists choose in their headlines at the same time.

6 to obtain a point estimate for the difference in differences.

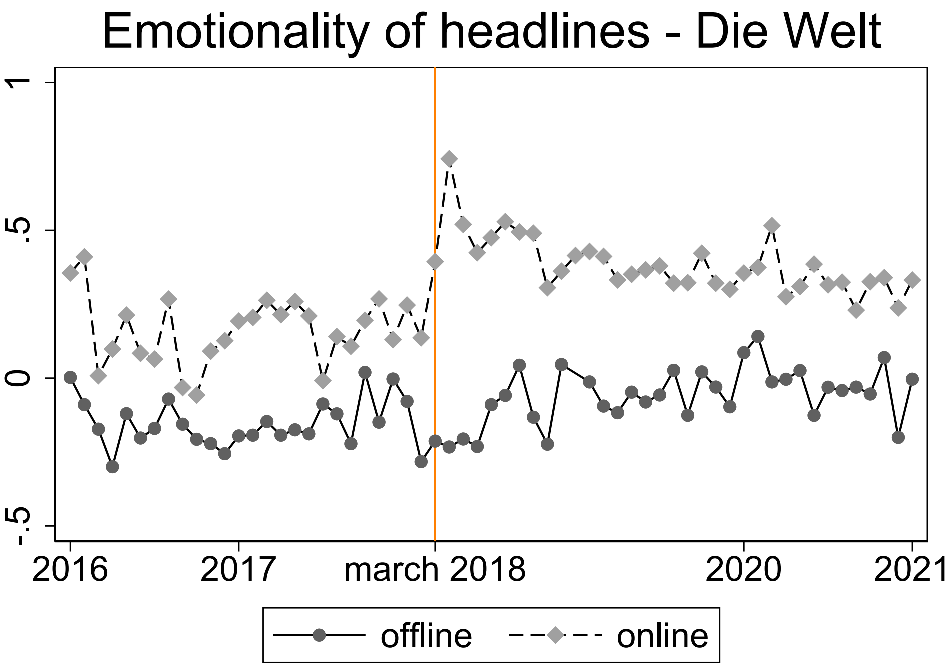
$$tonality_{it} = \beta_0 + \beta_1 online_i + \beta_2 D_i + \beta_3 online_i * D_i + \beta_5 time_t + \epsilon_{it} \quad (6)$$

$D_i$ is a dummy which is equal to one at and after the point in time when the automated news were introduced. $\hat{\beta_3}$ corresponds to the point estimate for the difference in differences. $time_t$ is a vector with dummies for each day to capture time fixed effects.

**Parallel trends** The analysis crucially depends on the validity of the parallel trends assumption. A visual inspection, an event study as well as placebo-tests for every month one year prior to the event suggest that the average emotionality of offline and online headlines of the agency content at *Die Welt* actually developed in parallel prior to the introduction of the automated finance news.[48] Also, as the articles are generated by a computer, it is not possible that this computer anticipated and influenced the emotionality of headlines prior to the introduction of the robo-news.[49]

Figure B22 illustrates the development of online and offline emotionality over time and allows for a visual inspection of the parallelity of the trends. The orange line indicates the month of the introduction of the automated news.

Figure B22: Emotionality of Welt Headlines over time (Agency only)



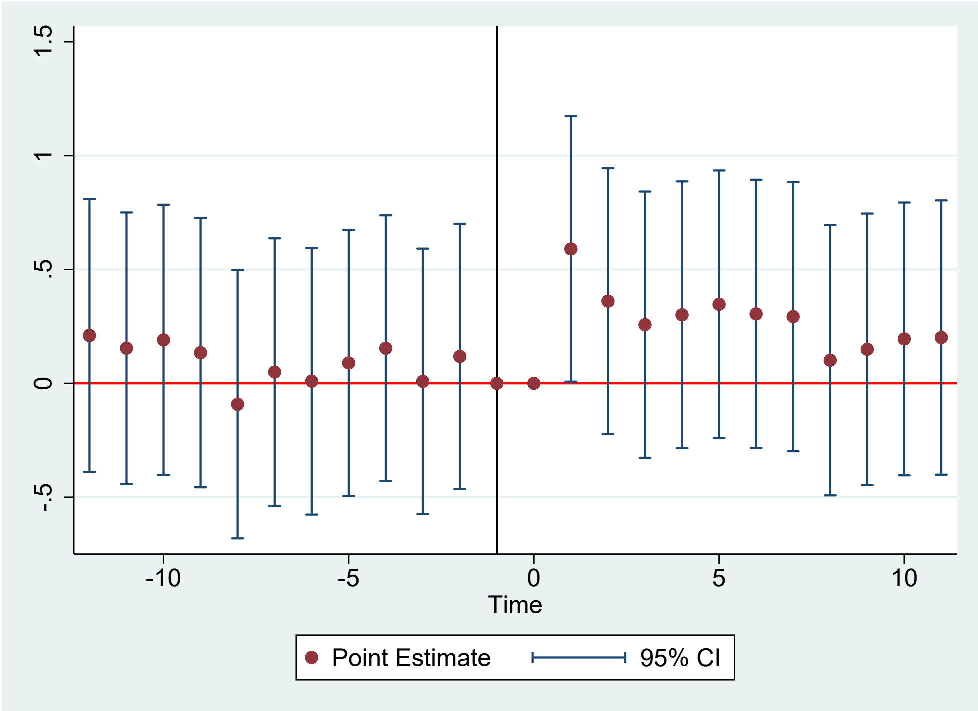

*Notes:* Figure B22 depicts the development of the emotionality of agency content headlines at *Die Welt* over time. The interval between each depicted datapoint is one month. Values for the offline headlines are illustrated by a full line. The online values are depicted with a dotted line. The orange line indicates the date on which the automated finance news were introduced.

Figure B23 is an event study graph, which shows that none of the periods before the event is statistically significantly different from zero, while all coefficients for the points in time after the event are positive and at least one of them is statistically significant on the 5 percent level.

[48] This is not true for sentiment. I therefore focus only on average emotionality here.

[49] As the articles were published in a very hidden section and did not target the entire readership, it is furthermore very unlikely that human reporters were aware of them before implementation or even changed the tonality in their reports prior to their introduction because of them.

Figure B23: Event Study - Automated News at Welt Online




*Notes:* Figure B23 depicts event study estimations of the emotionality of agency content headlines at *Die Welt* before and after the introduction of automated finance news. The interval between each depicted datapoint is one month and the total time frame regarded is one year before and one year after the event.

Figures B22 and B23 suggest that headlines of online articles became on average more emotional after the introduction of the automated reports. The gap in emotionality between online and offline headlines seems to widen and remains higher for several years. Running the estimation from equation 6 with dummies for "fake events" for each month one year prior to the actual event gives further support for the validitiy of the parallel trends assumption. Out of the 12 placebo regressions only one gives an estimate that is statistically significant (on the 10 percent level).

**Main finding** Running the estimation from equation 6 reveals that automated news increase the average emotionality of online headlines by 0.23 standard deviations (p=0.001). This corresponds to a headline being 13 percent more likely to be classified as emotional. The detailed regression results are available in column 1 of Table B14.

**Robustness: Alternative depended variables** I assess the robustness of the finding by using the classifications of the other sentiment classifiers as described in Appendix A.1 as alternative outcomes. The finding reproduces qualitatively with all other classification methodologies and the coefficient is in three out of the four cases statistically significantly different from zero on conventional levels. The size of the effect ranges from 0.07 to 0.46 standard deviations. Detailed results are provided in columns 2 to 5 of Table B14.

**Robustness: Time trends** A potential concern is that the result is driven by a time-trend in the online agency content that is captured by DiD-coefficient. To account for that I control for online/offline-specific linear time trends in an additional regression. Adding these controls does not change the result qualitatively. The coefficient becomes a bit bigger (from 0.25 to 0.32 standard deviations), but slightly less statistically significant (from p=0.001 to p=0.071). Details are available in column 6 of Table B14.

Table B14: OLS Estimates - Emotionality of Welt Headlines

| | tuned roberta (1) | sentiWS (2) | LM (3) | vader (4) | roberta (5) | tuned roberta (6) |
|---|---|---|---|---|---|---|
| online | 0.6731*** | 0.4887*** | 0.4691*** | 0.6963*** | 0.8238*** | 0.6634*** |
| | (0.0488) | (0.0444) | (0.0457) | (0.0479) | (0.0495) | (0.1056) |
| after | 0.3316 | -0.0776 | -0.3089 | -0.1777 | 0.2383 | 0.7214 |
| | (0.2909) | (0.3480) | (0.3347) | (0.2845) | (0.2695) | (0.5745) |
| online * after | 0.2506*** | 0.1660** | 0.2622*** | 0.4647*** | 0.0757 | 0.3159* |
| | (0.0730) | (0.0668) | (0.0711) | (0.0695) | (0.0743) | (0.1779) |
| time FE | yes | yes | yes | yes | yes | yes |
| time trends | no | no | no | no | no | yes |
| Constant | -0.8229*** | -0.1563 | -0.1257 | -0.5180** | -0.5718** | -0.9574*** |
| | (0.2453) | (0.2849) | (0.2757) | (0.2417) | (0.2330) | (0.2205) |
| $R^2$ | 0.1648 | 0.1265 | 0.1908 | 0.2381 | 0.1605 | 0.3745 |
| Observations | 11,278 | 11,278 | 11,278 | 11,278 | 11,278 | 1,818 |

*Notes:* Table B14 reports OLS estimates in standard deviations with robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

**Robustness: Alternative time windows** Furthermore, a concern could be that the result depends on the specific time frame used in the analysis. To explore this I repeat the analysis described in equation 6 but instead of a year I change the time frame to 18 or 6 months before and after the introduction of the automated news. The result is robust to both a expansion and limitation of the time frame. The estimate becomes a bit bigger when the time frame is smaller (0.42 standard deviations) and smaller when the time frame is bigger (0.11 standard deviations), but remains qualitatively the same and statistically significant on conventional significance levels.

**Discussion** The finding presented here, and especially the point estimate, should for two reasons be interpreted with caution. First, it is important to keep in mind that the comparison occurs on average sentiment and each article a newspaper publishes receives the same weight for these averages. Thus, the effect in the data at hand does not necessarily translate into a effect of comparable size for the average reader, because the automated news most likely reach only a very small fraction of readers with very specific interests, such as the day-to-day development of one specific stock. Second, the type of automated text generation analyzed here is only applicable to a small set of text types and business models. It should thus not be expected to be a trend that is augmenting to the entire news industry rapidly. Also, automated news in other contexts might have different effects on headline tonality and future news production robots might produce completely different headlines.

# C. Supplementary Material Readers' Reactions

## C.1. Sample Characteristics and Randomization Check

Table C15: Balance Table Readers

| | positive (1) | neutral (2) | negative (3) | t-test 1 VS. 2 (4) | t-test 1 VS. 3 (5) | t-test 2 VS. 3 (6) |
|---|---|---|---|---|---|---|
| age | 27.606 | 26.840 | 28.930 | 0.766 | -1.324 | -2.090** |
| | (0.562) | (0.567) | (0.800) | | | |
| male | 47.5% | 40.0% | 47.0% | 7.5% | 0.5% | 7.0% |
| *political preference* | | | | | | |
| SPD | 7.1% | 12.0% | 17.0% | -4.9% | -9.9%** | -5.0% |
| CDU/CSU | 12.1% | 11.0% | 7.0% | 1.1% | 5.1% | 4.0% |
| Die Grünen | 41.4% | 36.0% | 39.0% | 5.4% | 2.4% | -3.0% |
| FDP | 18.2% | 20.0% | 16.0% | -1.8% | 2.2% | 4.0% |
| AfD | 1.0% | 1.0% | 1.0% | 0% | 0% | 0% |
| Die Linke | 9.1% | 4.0% | 4.0% | 5.1% | 5.1% | 0% |
| other | 6.1% | 7.0% | 5.0% | 0.9% | 1.1% | 2.0% |
| wouldn't vote | 5.1% | 9.0% | 11.0% | -3.9% | -5.9% | -2.0% |
| phone use | 7.1% | 6.0% | 16.0% | 1.1% | -8.9%** | -10.0%** |
| ex-ante feeling | 12.323 | 12.430 | 12.440 | -0.107 | -0.117 | -0.010 |
| | (0.205) | (0.212) | (0.200) | | | |
| econ knowledge | 3.414 | 3.210 | 3.490 | 0.204 | -0.076 | -0.280** |
| | (0.095) | (0.099) | (0.099) | | | |
| finance knowledge | 2.909 | 2.690 | 2.960 | 0.219 | -0.051 | -0.270 |
| | (0.122) | (0.121) | (0.115) | | | |
| risk preference | 9.990 | 9.590 | 9.650 | 0.400 | 0.340 | -0.060 |
| | (0.225) | (0.222) | (0.222) | | | |
| Observations | 99 | 100 | 100 | | | |

*Notes:* political preference: percentage of people who would vote for a certain party if there were national elections on the next Sunday; phone use: percentage of participants answering the survey on their phone; ex-ante feeling: self-evaluation on 11-point Likert scale; econ knowledge: self-evaluation about knowledge about the economy on 5-point Likert scale; finance knowledge: self-evaluation about knowledge about finance on 5-point Likert scale risk preference: self-evaluation on 11-point Likert scale.
The value displayed for t-tests in column 4,5 and 6 are the differences in the means across the respective groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

## C.2. Results

### C.2.1. Emotional Reactions

Table C16: OLS Estimates - ATE on Emotions

| | Current Mood in SD | | Aggregate Affect in SD | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| positive | 0.2175* | 0.2029*** | 0.0428 | -0.0067 |
| | (0.1389) | (0.0655) | (0.1403) | (0.1137) |
| negative | -0.0390 | -0.0023 | 0.0276 | 0.0361 |
| | (0.1440) | (0.0702) | (0.1450) | (0.1249) |
| pre-feeling | | 0.4204*** | | 0.2500*** |
| | | (0.0175) | | (0.0271) |
| age | | 0.0024 | | 0.0043 |
| | | (0.0042) | | (0.0076) |
| phone | | 0.0027 | | 0.1442 |
| | | (0.0891) | | (0.1595) |
| male | | -0.0293 | | -0.0070 |
| | | (0.0647) | | (0.1171) |
| know econ | | 0.0161 | | 0.0905 |
| | | (0.0427) | | (0.0753) |
| know finance | | 0.0302 | | 0.0176 |
| | | (0.0415) | | (0.0608) |
| politics FE | no | yes | no | yes |
| Constant | -0.0598 | -5.5289*** | -0.0235 | -3.8670*** |
| | (0.1007) | (0.2777) | (0.1021) | (0.4738) |
| $R^2$ | 0.0128 | 0.7901 | 0.0003 | 0.3109 |
| Observations | 299 | 299 | 299 | 299 |

*Notes:* Table C16 reports OLS estimates with robust standard errors in parentheses. All outcomes are scaled in standard deviations. The group that was exposed to the neutral headline is always the reference group. Pre-feelings are the answers to the current mode question prior to randomization on an 11-point Likert scale. Age is scaled in years. Phone is a dummy for survey answering on a smartphone. Know econ and know finance are answers to the self-evaluation questions on economic and financial knowledge on a 5-point Likert scale. *** p<0.01, ** p<0.05, * p<0.1

### C.2.2. Reactions in Expectations

Table C17: OLS Estimates - ATE on GDP Expectations

| | Expectations for 2021 | | Expectations for 2022 | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| positive | 0.1959 | 0.1720 | 0.1821 | 0.1745 |
| | (0.1427) | (0.1399) | (0.1376) | (0.1435) |
| negative | 0.1549 | 0.0930 | 0.2781** | 0.2766* |
| | (0.1395) | (0.1449) | (0.1393) | (0.1427) |
| controls | no | yes | no | yes |
| Constant | -0.1170 | 06417 | -0.1531* | 0.2989 |
| | (0.0977) | (1.0388) | (0.0908) | (0.9668) |
| $R^2$ | 0.0072 | 0.0615 | 0.0133 | 0.0442 |
| Observations | 299 | 299 | 299 | 299 |

*Notes:* Table C17 reports OLS estimates with robust standard errors in parentheses. All outcomes are scaled in standard deviations. The group that was exposed to the neutral headline is always the reference group. Control variables are age, gender and political orientation of the participants as well as a dummy for phone-usage and self-reported measures for their feelings before the survey, knowledge on economics and knowledge on finance. *** p<0.01, ** p<0.05, * p<0.1

Table C18: OLS Estimates - ATE on DAX Expectations

| | Expectations for 2021 | | Expectations for 2022 | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| positive | -0.1212 | -0.1085 | -0.1276 | -0.1358 |
| | (0.1483) | (0.1463) | (0.1482) | (0.1473) |
| negative | -0.3101** | -0.2729** | -0.2703** | -0.2931** |
| | (0.1369) | (0.1345) | (.1373) | (0.1413) |
| controls | no | yes | no | yes |
| Constant | 0.1432 | 0.8887* | 0.1322 | 0.1802 |
| | (0.1069) | (0.5093) | (0.1065) | (0.4890) |
| $R^2$ | 0.0165 | 0.0648 | 0.0124 | 0.0628 |
| Observations | 299 | 299 | 299 | 299 |

*Notes:* Table C18 reports OLS estimates with robust standard errors in parentheses. All outcomes are scaled in standard deviations. The group that was exposed to the neutral headline is always the reference group. Control variables are age, gender and political orientation of the participants as well as a dummy for phone-usage and self-reported measures for their feelings before the survey, knowledge on economics and knowledge on finance. *** p<0.01, ** p<0.05, * p<0.1

# D. Experimental Instructions (English translation)

## D.1. Experiment with Journalists

*Participants received experimental instructions in German. Below I provide an English transcript. A dotted line indicates the next page of the survey. Explanatory comments (which were not displayed to participants) are indicated by blue, italic text.*

**Welcome!** Participation in this study takes **five minutes**. You can earn up to 15 Euro (but at least 5 Euro). The money will be transferred via PayPal within two weeks. All information provided here will be stored anonymously and used exclusively for research purposes. We will ask for your **PayPal address** at the end of the survey to be able to remunerate you. This data will be stored separately from the rest of your information and will be deleted after the payment is completed.

**Are you currently working as a journalist?**

O yes
O no

*Participants who answered "no" had to leave the survey.*

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

*This page was only shown to the treatment group.*

On the next page you will see a news article and three headlines. Your task is to choose the headline that you would most like to put above this article. The selected headline will be shown to a larger group of other participants (non-journalists). The more of those readers click on your selected headline, the more you will be paid.

**Attention: The amount of your payout in this study depends on how many readers click on the headline you have selected.**

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

*This page was only shown to the control group.*

On the next page you will see a news article and three headlines. Your task is to choose the headline that you would most like to put above this article. The selected headline will be shown to a larger group of other participants (non-journalists).

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Which of the following headlines would you most likely put above the article below?

O **Encouraging forecast: the German economy is expected to grow strongly again in 2022**
O **Forecast: This is how the German economy will develop in the near future**
O **Scary forecast: 2021 will be worse than expected for the German economy**

*In their autumn report, which published in October, leading economic research institutes expect gross domestic product to grow by 2.4 percent in 2021. In spring they had still expected an increase of 3.7 percent for this year.*

*The economic situation in Germany is still characterized by the corona pandemic, the report says. In the course of 2022, however, the German economy is expected to return to normal capacity. According to forecasts by the institutes, gross domestic product will increase by 4.8 percent in*

*2022. In their spring forecast, the institutes had only assumed an increase of 3.9 percent for the next year.*

Suppose you were allowed to formulate a headline yourself: What headline would you write?

*Participants could type their answer into a free text field. They had the option to see the text of the article again through clicking on a button.*

What aspects do you respect when choosing headlines for a given text? (multiple answers are possible)

O length of the headline
O comprehensibility
O active language (instead of passive expressions)
O usage of verbs
O usage of positive words
O usage of negative words
O usage of emotional words
O factual correctness
O most accurate description of the content of the text
O sparks curiosity
O topicality
O high degree of informativeness
O other

*When participants checked the box "other" an additional field opened in which they could type in a free text answer in order to specify their response.*

Which medium do you (mainly) work for?

*Participants could type their answer into a free text field.*

What is your main role in your journalistic work?
O writing/producing content
O edit content
O production of pictures, graphics and animations
O other

How many years have you been working in journalism?

*Participants could type their answer into a free text field but the answer type was restricted to be a positive number.*

How old are you?

*Participants could type their answer into a free text field but the answer type was restricted to be a positive number.*

What is your highest educational degree?

O no degree
O lower secondary school diploma

O higher secondary school diploma
O high-school diploma
O academic degree
O other

*There are two types of secondary schools in the German school system which refer to different abilities of the students, i.e. obtaining a degree from a lower secondary school ("Hauptschule") is easier than obtaining one from a higher seconary school ("Realschule").*

If there were federal elections next Sunday, which party would you vote for?

O CDU/CSU
O SPD
O Bündnis 90/Die Grünen
O FDP
O AfD
O Die Linke
O other
O I wouldn't vote.

In general, how willing or unwilling are you to take risks?

*Participants could answer the question on a 11-point Likert scale ranging form "not willing at all" to "very willing".*

How willing would you be to give up something that benefits you today in order to benefit more in the future?

*Participants could answer the question on a 11-point Likert scale ranging form "not willing at all" to "very willing".*

How willing would you be to give to a good cause without expecting anything in return?

*Participants could answer the question on a 11-point Likert scale ranging form "not willing at all" to "very willing".*

How much do you agree with the statement "I am narcissistic."? *(Note: the word "narcissistic" means selfish, self-focused, and vain)*

*Participants could answer the question on a 7-point Likert scale ranging form "don't agree at all" to "agree completly".*

How much do you agree with the statement "I suspect people have only the best of intentions."?

*Participants could answer the question on a 11-point Likert scale ranging form "don't agree at all" to "agree completly".*

*This page was only shown to the treatment group.*

Thank you!

By participating, you have earned 5 Euro plus an additional amount depending on the clicks on the headline you selected. In order to be able to pay you, we need your PayPal address. It will be stored separately and deleted after your remuneration. The payments will be made within the next few days.

*This page was only shown to the control group.*

Thank you!

By participating, you have earned 10 Euro. In order to be able to pay you, we need your PayPal address. It will be stored separately and deleted after your remuneration. The payments will be made within the next few days.

### D.2. Experiment with Readers

*Participants received experimental instructions in German. Below I provide an English transcript. A dotted line indicates the next page of the survey. Explanatory comments (which were not displayed to participants) are indicated by blue, italic text.*

**Welcome!**

At the **end of this study** you will be given a **withdrawal code**, please write it down. After your participation we will redirect you to the withdrawal form. You may have to enter this code there.

This participation in this study will take approximately 5 minutes.

O I consent to the above conditions.

.....................................................................................................................................................

In general, how do you feel right now?

*Participants could answer the question on a 11-point Likert scale ranging form "very bad" to "very good".*

.....................................................................................................................................................

How good or bad is your knowledge on current topics in the field of economics and business?

O very bad
O bad
O neither good nor bad
O good
O very good

.....................................................................................................................................................

How good or bad is your knowledge on current topics in the field of finance?

O very bad
O bad
O neither good nor bad
O good
O very good

.....................................................................................................................................................

In general, how willing or unwilling are you to take risks?

*Participants could answer the question on a 11-point Likert scale ranging form "not willing at all" to "very willing".*

.....................................................................................................................................................

On the next page we will show you the **headline** of a news article. **Please read this headline carefully.**

Attention: If you **click on the headline** you can read the **entire article**. **This will cost you 5 cents.** The contents of the article may help you make more informed decisions later in the study. You can earn up to one euro for the decisions you make later. Once you've clicked on the headline, you can open or close the article as many times as you want at no additional cost.

.....................................................................................................................................................

*This page was only displayed to participants in the positive treatment group.*

**Encouraging forecast: the German economy is expected to grow strongly again in 2022**

*When participants clicked on the headline the entire article from the journalist experiment was displayed. Otherwise they only saw the headline.*

*In their autumn report, which published in October, leading economic research institutes expect gross domestic product to grow by 2.4 percent in 2021. In spring they had still expected an increase of 3.7 percent for this year.*

*The economic situation in Germany is still characterized by the corona pandemic, the report says. In the course of 2022, however, the German economy is expected to return to normal capacity. According to forecasts by the institutes, gross domestic product will increase by 4.8 percent in 2022. In their spring forecast, the institutes had only assumed an increase of 3.9 percent for the next year.*

*This page was only displayed to participants in the neutral treatment group.*

**Forecast: This is how the German economy will develop in the near future**

*When participants clicked on the headline the entire article from the journalist experiment was displayed. Otherwise they only saw the headline.*

*In their autumn report, which published in October, leading economic research institutes expect gross domestic product to grow by 2.4 percent in 2021. In spring they had still expected an increase of 3.7 percent for this year.*

*The economic situation in Germany is still characterized by the corona pandemic, the report says. In the course of 2022, however, the German economy is expected to return to normal capacity. According to forecasts by the institutes, gross domestic product will increase by 4.8 percent in 2022. In their spring forecast, the institutes had only assumed an increase of 3.9 percent for the next year.*

*This page was only displayed to participants in the negative treatment group.*

**Scary forecast: 2021 will be worse than expected for the German economy**

*When participants clicked on the headline the entire article from the journalist experiment was displayed. Otherwise they only saw the headline.*

*In their autumn report, which published in October, leading economic research institutes expect gross domestic product to grow by 2.4 percent in 2021. In spring they had still expected an increase of 3.7 percent for this year.*

*The economic situation in Germany is still characterized by the corona pandemic, the report says. In the course of 2022, however, the German economy is expected to return to normal capacity. According to forecasts by the institutes, gross domestic product will increase by 4.8 percent in 2022. In their spring forecast, the institutes had only assumed an increase of 3.9 percent for the next year.*

In general, how do you feel right now?

*Participants could answer the question on a 11-point Likert scale ranging form "very bad" to "very good".*

The following words describe different feelings and sensations. Read every word, then indicate the intensity with which you experience the respective emotion at the moment. You can choose between five gradations.

| | not at all | a little | somewhat | much | very much |
|---|---|---|---|---|---|
| upset | ○ | ○ | ○ | ○ | ○ |
| attentive | ○ | ○ | ○ | ○ | ○ |
| afraid | ○ | ○ | ○ | ○ | ○ |
| determined | ○ | ○ | ○ | ○ | ○ |
| nervous | ○ | ○ | ○ | ○ | ○ |
| inspired | ○ | ○ | ○ | ○ | ○ |
| active | ○ | ○ | ○ | ○ | ○ |
| hostile | ○ | ○ | ○ | ○ | ○ |
| awake | ○ | ○ | ○ | ○ | ○ |
| ashamed | ○ | ○ | ○ | ○ | ○ |

You will now make four estimates and two decisions. One of these answers will be selected at random and will become relevant for your remuneration.

The following applies to the following questions: The closer the estimated value to the actual realized value, the higher your payout. You can earn up to one euro.

Note: Your payout is independent of what other participants value. **You should therefore enter the value that you consider most probable for each guess question.**

*Note: For the following two questions, enter your answer as a percentage. Example: If you expect a one percent increase, enter "1". If you expect a one percent drop, enter "-1".*

What do you expect: by what percentage will the German gross domestic product increase or decrease in the course of **2021** compared to 2020?

*Participants could answer the question in a free text field.*

What do you expect: by what percentage will the German gross domestic product increase or decrease in the course of **2022** compared to 2021?

*Participants could answer the question in a free text field.*

*Note: On December 3, 2021, the Dax closed at 15,169 points.*

What do you expect: With how many points will the Dax close on December 31, 2021 (i.e. at the end of **this** year)?

*Participants could answer the question in a free text field.*

What do you expect: With how many points will the Dax close on December 31, 2022 (i.e. the end of **next** year)?

*Participants could answer the question in a free text field.*

You can now invest all or part of 50 cents in the DAX. The invested money remains invested **until December 31, 2021**.

Your money (i.e. both the uninvested and the invested money) will be paid out on January 1st, 2022. The amount of uninvested money remains the same. The amount of money invested depends on the development of the DAX.

*Example: If the DAX rises by two percent and you invest 50 cents, you will be paid 51 cents. If it falls by two percent, you will receive 49 cents.*

How much of the 50 cents do you want to invest?

*Participants could answer the question in a free text field.*

You can now invest all or part of 50 cents in the DAX. The money remains invested **until December 31, 2022**.

How much of the 50 cents do you want to invest?

*Participants could answer the question in a free text field.*

How old are you?

*Participants could answer the question in a free text field, but answers were restricted to positive numeric values.*

What gender do you feel you belong to?

O male
O female
O diverse

If there were federal elections next Sunday, which party would you vote for?

O CDU/CSU
O SPD
O Bündnis 90/Die Grünen
O FDP
O AfD
O Die Linke
O other
O I wouldn't vote.

Thanks!

Your personal payout code is: XXX.

Please click on the arrow below to be redirected to the withdrawal form.