The Effects of Algorithmic Content Selection on User Engagement with News on Twitter

Erwan Dujeancourt, Marcel Garz

Jönköping International Business School, CEnSE, and MMTC

September 2022

Corresponding author: Marcel Garz, Jönköping International Business School, P.O Box 1026, SE-551 11 Jönköping. Email: marcel.garz@ju.se

Abstract

This paper investigates how Twitter's switch from a reverse-chronological timeline to algorithmic content selection in March 2016 influenced user engagement with tweets published by German newspapers. To mitigate concerns about omitted variables, we use the Facebook postings of these newspapers as a counter-factual. We find that the number of likes increased by 20% and the number of retweets by 15% within a span of 30 days after the switch. Importantly, our results indicate a rich-get-richer effect, implying that initially more popular outlets and news topics benefited the most. User engagement also increased more for sensationalist content than quality news stories.

Keywords: algorithm bias; Facebook; news quality; social media

Word count: 16,212

1. Introduction

Social media platforms use content selection algorithms[1] to help their users cope with the vast amount of information available online, typically by selecting the kind of content that users find most interesting and relevant. The basic workings of platforms' content selections algorithms are generally well understood. For instance, the extent of previous user engagement with certain content, the timeliness of the information, and the popularity among other users are often considered as important factors determining what is selected in users' feeds (e.g., DeVito, 2017). However, to protect their intellectual property and commercial interests, social platforms keep the exact configuration of their algorithms secret. This is problematic because algorithmic selection has been raising concerns that users are predominantly exposed to belief-confirming information, which could lead to ideological filter bubbles (e.g., Pariser, 2011; Bakshy, Messing, and Adamic, 2015). Content selection algorithms have also been blamed to promote hate speech, conspiracy theories, and fake news (e.g., Allcott, Gentzkow, and Yu, 2019; Müller and Schwarz, 2021; Cinelli et al., 2022), all of which could have negative repercussions for democratic processes and users' health. However, even if algorithms were made transparent and designed to avoid biases, it cannot be ruled out that there are unintended, potentially harmful consequences for society, such as the reinforcement of gender inequality (Lambrecht and Tucker, 2019; Cecere et al., 2020), or a reduction of diversity of consumption of entertainment content (e.g., Anderson et al., 2020; Holtz et al., 2020). Hence, it is crucial to better understand how content selection algorithms of social platforms work in practice.

In this paper, we investigate how Twitter's switch from a reverse-chronological timeline to algorithmic curation in March 2016 influenced user engagement with tweets published by professional media companies. Before this switch occurred, users were simply exposed to all tweets from accounts they followed, and the most recent tweets appeared on the top of their feed. With the algorithm, content has been automatically selected and ranked based on users' presumed interests and tastes. The switch is an ideal object of investigation because it allows us to compare the same platform with and without algorithmic curation. We focus on news story tweets published by professional media companies because of the dual function of these tweets. On the one hand, news story tweets serve as a marketing tool to create traffic to news outlets' websites, where clicks generate advertising revenue. On the other hand, these tweets may also influence users' political opinions and are thus relevant to democratic processes. Our choice to investigate user engagement – rather than exposure to content – is motivated by the fact that the former is more consequential than the latter, as it involves information processing and behavioral actions. User engagement is considered to be a more important "online currency" than exposure and often forms the basis for editorial decisions (e.g., Myllylahti, 2020).

[1] We use the terms "content selection algorithm" and "curation algorithm" synonymously.

Anecdotal evidence suggests that content selection algorithms increase user engagement by matching content to tastes (e.g., Farkas, 2016). Using a sample of 37 German newspapers, we investigate this mechanism along the most important dimensions of taste regarding journalistic content: sources, news topics, and quality. For that purpose, we collect data on the social media output of the newspapers around the date of the switch. To avoid biased results due to omitted variables, such as viral news stories or changes in the media agenda, we use the postings of the same outlets on Facebook as a counterfactual. In contrast to Twitter, Facebook has used algorithmic content selection since its inception.

Applying a difference-in-differences approach, our results confirm the expected increase in user engagement. In aggregate terms, we find that the average number of likes per tweet increased by approx. 20% within a time span of 30 days after Twitter's switch, while the number of retweets rose by ca. 15%. We obtain similar results when we use different estimation windows, negative binomial or OLS regression, and measure user engagement at the outlet-day level rather than for individual tweets/posts. Placebo exercises with fictional treatment dates add support that our results are not an artifact of unrelated developments. We also investigate various characteristics of tweets – such as the number of words, mean word length, and use of emotionally connotated words – and verify that the newspapers did not change their output in response to Twitter's switch. Hence, we are confident that our estimates measure the impact of algorithmic curation rather than potential adjustments in newspapers' behavior.

The finding that the switch to algorithmic curation increased user engagement at large is in line with common wisdom. However, it is not a priori clear whether Twitter's content selection algorithm bypasses editorial decisions of media companies by (de-)emphasizing certain content. As discussed below, social media algorithms may select content based on its overall popularity, individual user preferences, or both (Bozdag, 2013; Möller at al., 2018). If the popularity component dominates, we expect mainstream content to be more often selected into users' feeds, implying rich-get-richer dynamics. In contrast, if the algorithm emphasizes individual preferences, niche content would be more relevant (e.g., Anderson, 2006; Napoli, 2016).

We find that the increase in user engagement was larger for national than regional newspapers and for outlets with many Twitter followers. We also obtain evidence of rich-get-richer dynamics in terms of news topics. Those with much user engagement before the switch experienced a greater increase in likes and retweets than news topics with low pre-switch engagement. While our data do not allow us to assess how Twitter's switch affected consumption diversity within and across individual users, the rich-get-richer dynamics imply that users' feeds included less niche and more mainstream content after the switch. Given media companies' dependence on user engagement and website referrals, the introduction of algorithmic selection therefore likely created economic incentives to offer less diverse news and scale up the production of mainstream content, which could amplify tendencies towards company growth and ownership concentration (Noam, 2016).

We also find that tabloids benefited more than outlets committed to quality journalism, especially in terms of retweets. In line with that, the number of retweets increased more for shorter tweets and those including many exclamation marks and emotionally loaded words. These patterns are consistent with the interpretation that the switch pushed sensationalist news, to the detriment of journalistic quality. Sensationalist content often induces emotional reactions (e.g., Grabe, Zhou, and Barnett, 2001; Uribe and Gunter, 2007), which prompts many users to retweet the content due to the psychological need to share strong feelings with others (e.g., Rimé, 2009). To the extent that sensationalist headlines trigger negative emotions, such as anger and outrage, Twitter's content selection algorithm could be detrimental to civic discourses and users' well-being and mental health. Twitter's push for sensationalist news also seems problematic from a political perspective because lower levels of engagement with quality news could diminish voters' information and decrease their ability to make knowledgeable decisions at the ballot box (Prior, 2003). Hence, our results resonate with demands for a regulation of algorithms (e.g., Oremus, 2021), as recently fueled by revelations about Facebook triggering users' negative emotions despite knowing about the harmful consequences (Hagey and Horwitz, 2021).

2. Related Literature

Our findings contribute to research on algorithm bias (e.g., Flyverbom, Koed Madsen, and Rasche, 2017; Cowgill and Tucker, 2020; Cecere, Corrocher, and Jean, 2021), especially in the context of popularity-based rankings. It has long been argued that these kinds of rankings create rich-get-richer dynamics. In the case of search engines, for instance, the most popular websites are consistently shown on the top, which further increases the popularity of these pages (Mowshowitz and Kawaguchi, 2002; Cho and Roy, 2004). This process implies a bias because popularity does not necessarily equate to relevance and quality. Most news aggregators also rely on overall story popularity, which may increase consumption of mainstream content (Napoli, 2016). Popularity bias can be expected to be particularly pronounced in the context of social platforms because their users typically do not search for information but rely on the automatic selection of content in their feeds (Fortunado et al., 2006; Nikolov et al., 2018). Shmargad and Klar (2020) provide experimental evidence that algorithmic ranking of news articles by popularity causes users to conform to the most dominant viewpoints. We confirm the findings of these studies but do not restrict our analysis to the popularity aspect of algorithmic selection and investigate the heterogeneity of effects along further dimensions.

In that sense, our research is related to a study by Claussen, Peukert, and Sen (2021), who evaluate the heterogeneous effects of a recommendation system for online news on clicks. While their study evaluates differences within one news website, we analyze the differential effects of algorithmic curation across a number of outlets. In addition, their study compares the performance of a selection algorithm with news recommendations of human editors, whereas our results speak to the effects of adding algorithmic curation to human selection by news consumers.

Our findings are also closely related to a recent field experiment conducted by Twitter affiliates (Huszár et al., 2022). The authors study user exposure to tweets with political content for a large sample of users in seven countries. They find that algorithmic curation amplifies mainstream right-wing content more than mainstream left-wing content, compared to reverse-chronological selection. They also show that content from politically extreme sources does not receive more algorithmic amplification than content from moderate sources. Our study complements their results by investigating user engagement rather than exposure, and by providing evidence compiled by Twitter-independent researchers. In addition, our analysis is not restricted to political news but covers the full range of news topics.

In addition, we complement studies that investigate the determinants of user engagement with social media content. De Vries, Gensler, and Leeflang (2012) and Lee, Hosanagar, and Nair (2018) investigate characteristics of advertising messages and find that interactive posts and posts involving humor and emotion are associated with more user engagement. According to Trilling, Tolochko, and Burscher (2017), classical news values predict how often news stories are shared on Facebook and Twitter. Similarly, Garz, Sörensen, and Stone (2020) present evidence that Facebook users are more likely to engage with ideologically like-minded content than counter-attitudinal posts. Metallo et al. (2020) find that user engagement with content provided by Italian and Spanish municipalities is higher when users have time to be attentive. Cavusoglu et al. (2016) and Heimbach and Hinz (2018) study the role of platform design in user engagement. Their results illustrate that small changes made by platforms, such as adjustments to privacy settings or button design, can have substantial implications for businesses using these platforms. Our findings provide another example of the power of platform policy and design, resulting from the configuration of their selection algorithms.

Finally, our paper contributes to research investigating the impact of social platforms on the business strategies of news publishers. A central question of this literature is whether social media substitutes for or induces online news consumption (e.g., Hong, 2012). For example, Sismeiro and Mahmood (2018) find that news stories shared on Facebook increase traffic directed to news outlets' websites, while Aral and Zhao (2019) show that news sharing increases page views of the New York Times via regional spillovers. Others investigate the impact of social media metrics on editorial and journalistic decisions (e.g., Myllylahti, 2020; Peterson-Salahuddin and Diakopoulos, 2020; Cagé, Hervé, and Mayozer, 2020). Garz and Szucs (2022) investigate how the configuration of Facebook's news feed algorithm affects the long-run supply of news stories on the platform, finding that media outlets act in accordance with the incentives created by Facebook. Rather than studying the general equilibrium effects, our research design focuses on short-run outcomes, which allows us to isolate changes caused by algorithmic content selection from possible adjustments made by news publishers.

3. Reverse-chronological timeline vs. algorithmic content selection on Twitter

When Twitter was launched in July 2006, the user was exposed to content through a timeline of tweets. This timeline simply listed all content shared by the user's network of followed accounts, starting with the most recent tweets posted at the top of the user's timeline. Therefore, exposure to content was exclusively based on the decisions of the consumer about which accounts to follow. In 2010, the platform introduced "Suggestions for you", an algorithm-based feature designed to assist users in finding relevant accounts they could follow. However, the timeline of tweets still exposed users to all of their "subscribed" content in reverse chronological order. With a growing number of users and an increasing amount of available content, Twitter decided to start testing a content selection algorithm in January 2015, using the label "While you were away". Initially, this feature was available only for iOS users. Android and desktop versions were rolled out in February and May 2015, respectively (Rosania, 2015). Importantly, the algorithmic content selection was an opt-in feature at that time; interested users had to access their Twitter settings and consciously activate it.

This policy was changed on March 17, 2016, when the platform modified the default option, forcing users to opt out of the "Best tweets first" feature if they still wanted to use a reverse-chronological rather than an algorithmically curated timeline (Kamen, 2016; Read, 2016; Schroeder, 2016). Both the opt-in and opt-out policies allowed users to choose their preferred version of the timeline. In many contexts, however, status quo bias causes consumers to stick with the default option (Samuelson and Zeckhauser, 1988; Kahneman, Knetsch, and Thaler, 1991). Several weeks after switching the default option, Twitter confirmed that only 2% of its users decided to opt out of algorithmic curation (Perez, 2016). Therefore, we consider March 17, 2016, the date when the vast majority of users were switched from reverse-chronological exposure to algorithmic curation.[2]

We assume that before March 17, 2016, exposure to content on Twitter was primarily driven by user decisions. For instance, a politically conservative consumer might have followed a conservative news outlet, whereas an apolitical user might have followed an entertainment- or sports-centered outlet. With Twitter's

[2] Given that a small fraction of users was exposed to algorithmic curation before March 17, 2016, the regression results presented in Section 6 could slightly underestimate the true magnitude of Twitter's switch. The same applies to unannounced A/B tests of certain features of the algorithm, which Twitter may have conducted prior to the global rollout.

chronological timeline, both users might have shown some level of engagement with the tweets of their chosen news outlet. However, since the chronological timeline listed every single tweet of a "subscribed" news account, users likely encountered a fraction of irrelevant or counter-attitudinal tweets. For example, the conservative news outlet might have occasionally reported news that contradicted the beliefs of the conservative consumer. As consumers might do when reading a printed newspaper, users likely skipped tweets that they did not enjoy or find relevant; for instance, due to cognitive dissonance (e.g., Zaller, 1992; Stroud, 2011). Eventually, users might have unfollowed the outlet if the fraction of unwanted tweets was too large. Consequently, there are limits to how much user engagement a chronologically organized timeline can create.

The algorithmically curated timeline that became the new default on March 17, 2016, has been designed to decrease the chances that users are exposed to irrelevant or counter-attitudinal content. As stated by Twitter, the algorithm selects tweets based on users' location and interests, the kinds of tweets and accounts they usually interact with, and the extent of engagement with tweets by other users (Farkas, 2016). In technical terms, the selection process therefore involves the following elements: (i) ranking of content according to the popularity among all users, (ii) semantic filtering based on past and current preferences of the individual user, and (iii) collaborative filtering based on the preferences of followers (Bozdag, 2013; Möller at al., 2018). If the algorithm worked as intended, a conservative user who "subscribed" to a conservative news outlet would still be exposed to many tweets of that outlet, but tweets of the outlet that contradicted the user's beliefs or were not of interest would be omitted from the timeline. In this case, we should observe an increase in user engagement after the switch because of a higher level of exposure to preferred content.

4. Data

4.1 Sample

Our sample of news outlets consists of 37 German newspapers that have traditionally pursued a two-sided revenue model of advertising and audience sales. We focus on a single country to avoid confounding effects related to cultural and institutional differences in news production and consumption. Our choice to study newspapers in Germany is motivated by the availability of complementary data and to gather evidence on social media from a context other than the much-studied US case. The selection of outlets includes the most read mainstream newspaper brands. This kind of sample allows us to discuss the implications of our findings for society at large, rather than insights that apply to niche topics or politically extreme opinions. We apply the following criteria to select outlets for inclusion in the sample: First, the newspaper needs to be included in the Genios database, Germany's largest, publicly accessible newspaper archive (https://www.genios.de/), which ensures that the outlet is an established brand in the German media landscape. Second, the outlet must have had both an official Twitter account and an official Facebook Page since at least 2013, which we evaluate by checking the "About" information on the newspapers' Twitter accounts and Facebook Pages.[3] With this criterion, we verify that the outlet had enough time to establish its brand on social media before Twitter introduced the curation algorithm. The resulting sample includes the national dailies *Frankfurter Allgemeine Zeitung, Frankfurter Rundschau*, *Handelsblatt*, *Süddeutsche Zeitung*, *Die Tageszeitung*, *Die Welt*, and *Welt Kompakt*; the national weeklies *Focus*, *Der Spiegel*, and *Die Zeit*; and 27 local newspapers. These outlets accounted for approximately 33% of the market share of online news in March 2016, if measured in terms of page impressions registered by the German Audit Bureau of Circulation – IVW. A list of the newspapers is provided in Table A1 in the Online Appendix.

We collect data on all 56,621 tweets and 44,571 Facebook posts of the outlets within a window of 30 days before and 30 days after March 16, 2016.[4] For that purpose, we scrape and parse historical information provided by the Twitter Search browser application. For each tweet, we collect the date and time of publication, the number of likes and retweets, and the message text.[5] We retrieve the same information for the Facebook posts using the Facebook Graph Application Programming Interface (API). The data collection took place between 2017 and 2019, which guarantees that we obtained the "final" engagement metrics.[6]

In addition, we collect the number of followers on Twitter before the switch to capture the outlets' initial popularity on the platform. We also classify the outlets by their overall journalistic approach. German newspapers can be divided into tabloid journalism (i.e., focusing on emotional content, sensationalism, and gossip) and quality-oriented broadsheet journalism. Our sample includes three known tabloids (*B.Z.*, *Express*, and *Hamburger Morgenpost*), while the remaining outlets fall in the category of broadsheet journalism. We use a binary variable to capture this distinction. We complement this information with a continuous, survey-based measure of news quality, as provided by Wellbrock (2011).[7] This measure is available for 26 outlets in our sample. It ranges from 4.67 (lowest observed quality) to 8.38 (highest quality).

[3] We choose this year because most of the German news industry adopted social media for the dissemination of news stories between 2010 and 2012 (Cision, 2013).
[4] We discard the option to obtain information on individual users because collecting those data would be a violation of the platforms' site policies and challenging from a research ethics perspective, given data protection regulations in the EU.
[5] Restrictions by Twitter of both its search browser interface and its API have prevented us from collecting consistent information about the number of comments on tweets. We do not believe that this limitation poses a major shortcoming, because comments capture a qualitatively different type of engagement than likes and retweets/shares (Yoon et al., 2018). Specifically, commenting often takes place in the context of controversial issues and thus counter-attitudinal content.
[6] Typically, 99.9% of liking and sharing takes place within 14 days of content creation; see Lee, Hosanagar, and Nair (2018).
[7] We assume that the survey data provide a valid measure of outlets' journalistic quality in March 2016, when Twitter switched to algorithmic curation, given the long-standing professional structures in the German news industry (Hanitzsch et al., 2011).

Finally, we construct a binary variable to distinguish between national and regional outlets, based on the newspapers' self-descriptions. See Table A2 for summary statistics of these outlet-level variables.

4.2 User engagement

Our dependent variables are the number of likes and retweets per tweet in the case of Twitter and the number of likes and shares per post in the case of Facebook. During our investigation period, the mean number of likes per tweet was 4.40 (SD = 17.74) and that of was retweets 4.11 (SD = 23.71). These numbers were substantially higher on Facebook, where we observe an average number of likes per post of 106.66 (SD = 997.67) and average shares of 30.91 (SD = 334.05); see Table A3. It could be argued that the number of likes per news story is a more informative unit of observation, as the news outlets might have reacted to Twitter's switch to algorithmic curation by increasing their daily number of tweets (e.g., to conduct more A/B testing), which would deflate the number of likes per tweet. Our data indicate that the daily average number of tweets increased by a negligible amount after the switch (see Figure A4, Panel A), which allows us to rule out any significant mechanical depression of per-tweet user interactions. In addition, robustness checks confirm that our results hold when we use the outlets' daily total number of likes as a dependent variable; see Table A10. Importantly, we refrain from investigating user engagement at the story level because A/B testing, the common use of URL shortening, and deleted content make it difficult to accurately identify unique stories. Furthermore, we cannot rule out that the news outlets changed the number of stories shared via Twitter, as in the case of the number of tweets.

4.3 Text characteristics and topic modeling

We assess the quality of individual tweets and posts based on their text characteristics. As previous studies show, messages produced to drive clicks are more likely to start with question words, include question or exclamation marks, and contain emotionally connotated words. In contrast, quality journalism is often characterized by relatively lengthy texts and longer words, due to a higher information content and more sophisticated language (e.g., Lacy and Rosenstiel, 2015; Choi, Shin, and Kang, 2021; Lischka and Garz, 2021). Hence, we construct variables that capture the number of words, the mean word length, a dummy variable to capture if a tweet/post starts with an interrogative (i.e., what, why, when, who, which, or how), and the number of question and exclamation marks. We use the SentiWS dictionary (Remus, Quasthoff, and Heyer, 2010) to compute the shares of negatively and positively connotated words per tweet/post.

We conduct topic modeling to determine the dominant news category of the tweets and posts. Traditional approaches – such as Latent Dirichlet Allocation (LDA; see Blei, Ng, and Jordan, 2003) – do not perform well when they are applied to short texts, due to the data sparsity of those texts. This limitation is particularly relevant in our case, as the average tweet consists of approximately 12 words. In fact, we fail to obtain meaningful results when we apply LDA to the data at hand. Instead, we estimate biterm topic models (Yan et al., 2013), which are designed to identify topics in short texts, as commonly found online and on social media. This approach relies on co-occurrences of words in the entire corpus rather than patterns within individual documents.

We lemmatize the tweets and posts in our sample (Straka and Straková, 2017), extract all observed biterms (i.e., pairs of words), and model the probability that a given biterm is drawn from a specific topic. In contrast to LDA, the optimal number of topics cannot be determined by calculating perplexity scores via held-out tests. The statistical performance of different biterm topic models can instead be assessed by comparing the sum of log-likelihoods over biterms in a given model (Yan et al., 2013). Figure A1 plots this metric for models with different numbers of topics. The figure indicates that the best statistical fit is achieved by models that assume more than 100 topics, in which case the results are too complex to offer any meaningful interpretation. However, for models that assume more than 20 topics, the gains in statistical fit are marginal. That is, the slope of the curve is basically flat in that area, which suggests that there may not be much value in considering models that assume a number of topics larger than that.

We therefore evaluate the interpretability of models within a range of 15 to 25 topics, by looking at the top terms associated with the topics. Specifically, we assess whether the resulting topics can be unambiguously labeled in terms of typical news categories. We find that a model with k = 18 topics offers the best interpretability. With this model, we identify news topics on politics, the economy, business, foreign issues, sports, weather, panorama, technology, and traffic accidents/reports, as shown in Table A4. In addition, there are three crime-related topics, covering police investigations, court proceedings, and corporate crime. The model also identifies two topics addressing the 2015/16 European migrant crisis (a main news topic at the time), one related to political aspects and another one covering crimes committed by refugees. One topic refers to tweets/posts that include calls to action – as often found on social media – such as “tell us your opinion” and “please like this picture”. Finally, the model groups together tweets/posts addressing outlets’ own matters (e.g., server issues, editorial comments) as well as those advertising certain highlights of the print editions. With a lower number of topics, models tend to group conceptually distinct news categories together, such as sports and politics. In contrast, it is difficult to find labels for certain topics if we choose a larger number than k = 18 topics.

We define the dominant topic of each tweet and post as the one with the highest topic weight. As shown in Table A4, tweets and posts related to “sports” have the highest share in the sample (8.83%), whereas the category “highlights of the print edition” accounts for the lowest share (2.21%). We measure the popularity of the topics during the 30 days before Twitter’s switch by computing the mean number of likes over tweets belonging to the same news topic. Tweets discussing “crime committed by refugees” on average received the highest number of pre-switch likes (4.08) and “business” news had the lowest number (1.96).

4.4 Matching of tweets and posts

In some cases, news outlets publish the same news story on Twitter and Facebook. In other cases, a story is only published on Twitter or only on Facebook. Identifying “cross-platform twins” allows us to split our sample into matched and unmatched tweets/posts, which helps to address concerns about the validity of Facebook as counterfactual. Using term-frequency vectors, we compute the cosine similarity between all tweets and posts of a newspaper on the same day. The cosine similarity is a standard measure in computational linguistics to evaluate the resemblance of texts (e.g., Jurafsky and Martin, 2008).[8] It varies between 0 and 1, with 0 indicating no resemblance between two texts and 1 indicating a perfect match. Thus, the approach accounts for the possibility that the same news story is published in different ways on Twitter and Facebook. For instance, a news story about Donald Trump wearing a mask could be published by the same news outlet as “Change of mind: Trump recommends face masks” on Twitter and perhaps as “Donald Trump advises the public to wear a mask” on Facebook; see Figure A2 for real examples of matched and unmatched tweets/posts. Our variable of interest is the maximum cosine similarity that a tweet yields among all Facebook posts on the same day by the same newspaper, and vice versa (i.e., the maximum cosine similarity that a Facebook post has among all tweets on the same day by the same newspaper). This variable has a sample mean of 0.31 (SD = 0.23), which suggests a small to moderate overlap between Twitter and Facebook content and is in line with previous research showing that outlets customize their tweets and posts for the different audiences on these platforms (Lischka and Garz, 2021). After the matching, differences between tweets and posts do not disappear but are reduced along some dimensions. For example, the matched tweets and posts tend to have the same average length and similar shares of negatively and positively connotated words but still differ in mean user engagement.

[8] We also consider matching tweets and posts using BERT (Devlin et al., 2018) and other techniques that account for word embeddings (e.g., Kenter and de Rijke, 2015) but spot checks suggest that cosine similarities between term-frequency vectors perform best with the data at hand. An alternative approach would be to match on the URLs linking the tweets and posts to the respective news story on the outlet’s website. However, this approach is not feasible in our context. Due to Twitter’s character limit, the majority of tweets include shortened URLs (e.g., using Bitly and other services). In contrast, the outlets typically use long URLs in their Facebook posts because the platform offers a dedicated format to share links. More than half of the shortened URLs could not be converted to long URLs at the time of collecting the data – because the relevant links do not exist anymore – which makes it impossible to compare URLs across platforms in many cases.

5. Estimation

Estimating the effect of the Twitter's switch on user engagement is difficult because in most cases, the algorithm and the user both strive for exposure to relevant and belief-confirming information. Therefore, the effects on engagement are observationally equivalent. Comparing user engagement before and after Twitter's implementation of algorithmic content selection allows us to tackle the issue of observational equivalence. However, a plain before-after comparison could be problematic because special news events, seasonal effects, or unobserved factors could coincide with the time of the switch. For that reason, we compare the change in user engagement on Twitter with the user engagement that we observe for the same news outlets on Facebook. This research design can be implemented as a difference-in-differences regression of the following form:

$$engagement_{i,n,p,t} = \alpha_1 + \alpha_2 TW_p + \alpha_3 After_t + \alpha_4 TW_p \times After_t + \alpha_5 X_{n,p,t} + \varepsilon_{i,n,p,t} \qquad (2)$$

where $engagement$ is the number of likes or retweets/shares pertaining to tweet/post i, published by outlet n on platform p (i.e., Twitter or Facebook) and date t. TW is a dummy that takes the value 1 for Twitter and 0 for Facebook. $After$ equals 1 for observations strictly after March 16, the date when Twitter switched to algorithmic content selection as the default option. The coefficient α_4 on the interaction between TW and $After$ measures the treatment effect of the switch on user engagement. The vector X includes day and outlet-platform fixed effects and an outlet-platform specific trend polynomial of order 3. Day fixed effects account for the possibility that user engagement varies over the course of the week or month, whereas outlet-platform fixed effects control for time-invariant differences across Twitter accounts and Facebook Pages. The outlet-platform specific trend polynomial helps to address concerns that the estimates are driven by differential time trends in user engagement between Twitter and Facebook, rather than Twitter's switch to algorithmic curation. We do not add further controls, such as tweet/post characteristics, because they could be colliders between TW and $TW \times After$ and thus lead to a biased estimate of α_4. Equation (2) is estimated as a negative binomial model because the engagement metrics are count variables that exhibit overdispersion. Consequently, all regression coefficients are reported as incidence rate ratios, which can be easily interpreted as percentage changes in our case. Standard errors refer to these ratios and are clustered at the outlet-platform level to account for heteroskedasticity and correlation within outlets' Twitter accounts and Facebook Pages.

Using Facebook as a counterfactual helps us to alleviate concerns that changes in engagement were driven by common social media effects or global shocks to the news agenda. However, the outlets' news stories on Facebook can serve as a meaningful counterfactual only if the determinants of user engagement on that platform remained unchanged. Facebook has been using algorithmic content selection for its news feed since the public launch of the platform in 2006, but the algorithm has often been changed since then. Some updates of the news feed algorithm have explicitly targeted professional news stories. For instance, in December 2013, Facebook announced that it would prioritize high-quality news (Garz and Szucs, 2022); in August 2014 and again in August 2016, the platform changed its algorithm with a design to curb the proliferation of clickbait (Peysakhovich and Hendrix, 2016).

To obtain a valid counterfactual, we therefore restrict the estimation of Equation (2) to tweets and posts published within a few days around Twitter's switch to algorithmic curation (i.e., +/- 7 days, +/- 14 days, +/- 30 days). This restriction minimizes the risk that we include (potentially unannounced) changes to Facebook's news feed algorithm that might have affected the user engagement with the postings of the outlets in our sample. We also check whether any announced updates of the news feed algorithm took place close to the relevant date. The closest changes that might be important here occurred 22 days before and 36 days after Twitter's switch to algorithmic curation.[9] Thus, a very short estimation window (e.g., +/- 7 days) gives us confidence that our results are not biased by changes in Facebook's news feed algorithm, whereas a broad window (e.g., +/- 30 days) allows us to a get sense of the effects over a longer time period. In addition, a short estimation window decreases the risk of confounding due to potential reactions of the newspapers to the switch.

Figure 1 shows the development of the average number of likes and retweets/shares on Twitter and Facebook at the relevant time. The figure reveals that likes followed parallel trends on the two platforms before Twitter's switch to algorithmic curation. The same can be observed for retweets/shares, although the increase in retweets four to five days before the switch is not necessarily paralleled by the development of shares. After the relevant date, we observe a higher level of user engagement on Twitter but not on Facebook. The fact that user engagement on Facebook was not affected by the switch supports our claim that posts published on this platform are a valid counterfactual. We discuss further threats to the validity of the counterfactual in the next section.

6. Results

6.1 Aggregate effect

Table 1 summarizes the regression results for different time windows around the switch. Throughout, the coefficient on the *Twitter* dummy is significantly smaller than 1. Hence, the number of likes and retweets was generally lower on Twitter than on Facebook during the pre-algorithm period. The coefficient on the

[9] On February 24, 2016, Facebook announced the global rollout of Reactions (Krug, 2016). On April 21, 2016, the platform announced an update that would allow the news feed algorithm to account for results of user surveys (Blank & Xu, 2016).

After dummy is insignificant in all models, which confirms that user engagement on Facebook did not change after March 16. However, as the interaction between both variables indicates, the number of likes increased on Twitter after the platform introduced its curation algorithm. The effect is significant at the 1% level, regardless of the length of the estimation window. The coefficient ranges between 1.185 (+/- 14 days) and 1.208 (+/- 30 days), which implies an increase in Twitter likes between 18.5% and 20.8% after the switch to algorithmic curation. For retweets/shares, the coefficient on the *Twitter* × *after* interaction indicates increases by 7.1% (+/- 14 days) and 15.1% (+/- 30 days). However, the standard errors are somewhat larger than in the case of likes, which implies that the effect is only significant when using an estimation window of +/- 30 days. Likes and retweets are correlated with each other, but we do not necessarily expect symmetric results for both types of engagement, as they serve different functions for users, which we discuss in Section 6.3. The estimated effect sizes are similar to the effect sizes that Claussen, Peukert, and Sen (2021) find in their comparison of algorithmic and human editor content selection of a German news website. They find that algorithmic selection on average increases clicks by 4% and a five times larger effect for heavy news users.

To verify that the changes in user engagement were actually driven by Twitter's switch, we regress the engagement metrics on interactions between the *Twitter* dummy and lags and leads of the time of the switch. As Figure A3 in the Online Appendix shows, there are no significant differences in likes and retweets/shares between Twitter and Facebook in the pre-switch period. In contrast, we obtain multiple significant coefficients that are larger than 1 in the post-switch period, which supports the interpretation that likes and retweets increased because of the switch.

A major concern of using Facebook as a counterfactual is that the audiences on Twitter and Facebook are different. For example, Twitter users tend to be better educated than users on Facebook and thus have a stronger preference for quality journalism (e.g., Cornia et al., 2018). It would be problematic for our research design if these differences imply that users react to algorithmic curation in different ways. In particular, it could be argued that Facebook users may not react at all to the algorithmic selection of news stories since they care mostly about their friends' photo posts and life events. We do not believe this concern poses a major threat in our context because user engagement with news story posts is largely driven by users that are followers of news outlets (Cornia et al., 2018). That is, our research design does not compare Twitter and Facebook users at large but audiences that use social media for news consumption. A related concern refers to multi-homing, that some users consume news on both Facebook and Twitter. The estimates in Table 1 could be biased if users refrain from engaging with a news story on Facebook because the Twitter algorithm already exposed them to the same story on Twitter (or the other way around).

Note that both concerns – bias due to user differences between audiences and bias due to multi-homing – present opposite ends of the same threat to our research design. It would be optimal to investigate the issue with data on individual users, but we do not have those data at our disposal. However, it is possible to assess if either type of confounding leads to a substantial bias in the estimates by splitting up the sample into matched and unmatched tweets/posts. Considering matched tweets and posts (i.e., those items that have a "cross-platform twin") helps us to alleviate concerns about bias due to differences between Twitter and Facebook users, because users engaging with the same content can be assumed to be rather similar. In contrast, a sample using unmatched tweets and posts (i.e., those items that only appeared on one of the two platforms) allows us to mitigate concerns about multi-homing. If algorithmic exposure to content on one platform affects user engagement on the other platform in a systematic way, we would expect different results when using the subsample of unmatched tweets and posts, compared to our baseline estimates.

Table A5 summarizes the results of this exercise, using the cosine similarity values of tweets and posts to define subsamples of matched and unmatched content. Regarding likes, we obtain estimates that are very similar to the full sample, with coefficient values that imply comparable effect sizes of around 20%. An exception are regressions with few observations (i.e., Panel A, Columns 1 and 2), which we attribute to the lower statistical power of these models. In the case of retweets/shares, the effect sizes are somewhat larger than in the baseline regressions, especially for the matched sample. Here we estimate an increase in retweets between 25.7% to 29.3%, compared to 7.1% to 15.1% in the overall sample. A possible explanation for this difference is that news stories published in similar form on both platforms could be generally more newsworthy and appealing to a broader audience than stories that are only posted on one of the two platforms. Hence, these results are compatible with the rich-get-richer effects discussed throughout the paper.

The estimates in Panel A are unlikely biased due to differences between Twitter and Facebook users, but these estimates could be entirely driven by multi-homing. In contrast, the results in Panel B are unlikely biased due to multi-homing, but they could be entirely driven by differences between audiences. Note, however, that both sources of confounding cannot simultaneously affect our results: If a large fraction of users engages in multi-homing on Twitter and Facebook, there cannot be large differences between audiences. If there are large differences between audiences, there cannot be much multi-homing. Given this constellation, jointly interpreting the results in Panels A and B gives us confidence that neither factor acts as a strong confounder in the effect of Twitter's switch on user engagement. If differences between audiences induced estimation bias, we would obtain qualitatively different results in Panel A. If multi-homing led to estimation bias, we would see qualitatively different results in Panel B. In other words, knowing the Panel A results offers reassurance that the Panel B estimates are unlikely subject to hidden bias stemming from differences between audiences, whereas the Panel B results allow us to rule out hidden bias in Panel A due to multi-homing.

To further strengthen this argumentation, we apply the method outlined by Cinelli and Hazlett (2020) and assess the sensitivity of the estimated treatment effects to unobserved confounding. Specifically, Table A5 reports the *partial R^2 of the treatment* with the outcome and the *robustness value* for the treatment point estimate. In regressions with a statistically significant treatment effect, the *partial R^2 of the treatment* suggests that Twitter's switch explains between 4.7% (Panel B, Column 6) and 20.6% (Panel C, Column 3) of the variation in likes and retweets. These magnitudes are particularly informative in relation to the *robustness value*, which is also expressed in terms of partial R^2. This metric captures how strongly one or more confounding factors would have to be associated both with the treatment and the outcome to drive the estimated treatment effect to zero. In regressions with a statistically significant treatment effect, the robustness value ranges between 19.9% (Panel B, Column 6) and 39.6% (Panel C, Column 3), which is approximately two to three times as much as the partial R^2 of the treatment with the engagement metrics. Hence, confounding factors would not only have to explain both user engagement and selection of users on Twitter vs. Facebook (i.e., the relevant part of the residual variance of the treatment) but do so in a much more powerful way than Twitter's switch for the estimated treatment effects to disappear. It seems unlikely that multi-homing or differences between users could induce changes in the engagement metrics – at the time of the switch – that are at least twice as strong as the actual effects of the switch.

In Table A6, we verify that our results do not depend on the estimation method. That is, we re-estimate Equation (2) using ordinary least squares (OLS) – rather than negative binomial regression – and the logarithm of the number of likes and retweets/shares. These estimates indicate positive effects for all estimation windows. However, the coefficient values imply smaller effect sizes, ranging between 4.7% (+/- 14 days) and 7.9% (+/- 30 days) in case of likes and 1.7% (+/- 30 days) and 7.4% (+/- 7 days) for retweets/shares, which could be related to OLS producing biased estimates when count data are involved. Given that log transformations might be problematic when the value 0 appears, we also report OLS estimates when using the inverse hyperbolic sine transformation of the engagement metrics (Burbidge, Magee and Robb, 1988). The corresponding results in Table A7 are nearly identical to those in Table A6.

The placebo regressions shown in Tables A8 and A9 offer further support for our findings. In these regressions, we re-estimate our baseline model while using fictional treatment dates (i.e., 12 and 3 months before and 3 and 12 months after Twitter's switch). These specifications do not indicate any statistically significant effects. Thus, we can be confident that the results in Table 1 are not a statistical artifact and that our empirical approach does not pick up some unrelated trend.

Finally, we replicate our analysis with data at the outlet-platform-day level, using the daily sum of likes as the outcome variable. This exercise allows us to rule out the possibility that changes in the number of tweets after Twitter's switch distort our baseline estimates by deflating or inflating per-tweet engagement metrics. As Table A10 shows, the resulting estimates indicate changes that are very similar to the tweet-level data, with effect sizes as large as 16.9% for likes (+/- 7 days) and 16.1% for retweets/shares (+/- 30 days).

6.2 Alternative explanations

Our primary interpretation of the results presented in Table 1 is that Twitter's switch to algorithmic curation led to a boost in user engagement by exposing news consumers to more content that matched their tastes. This subsection discusses the plausibility of possible alternative explanations.

First, Twitter's switch could have caused users to be more active on the platform simply due to an increase in curiosity about the novel features. Indeed, the switch received widespread attention and criticism (Perez, 2016). However, if awareness and curiosity were the driving force, we would expect user engagement to briefly increase for only a few days. Once this curiosity ceased, user engagement should revert. This is not the case though. We observe an increase in likes for at least 30 days after the algorithm change (Table 1, Columns 3 and 6), which is not compatible with a temporary curiosity effect. Data on monthly active Twitter users do not support the curiosity explanation either (see Figure A6). If Twitter's switch had increased the attractiveness of the platform, one would expect a discontinuous increase in the number of users at that point, but the figure does not indicate any discontinuity in user growth.

Second, it could be argued that our results are driven by extended promotion activities of the news outlets. Twitter offers various tools for businesses to increase the reach and impact of tweets, such as "quick promotion", conversion tracking, and objective-based advertising programs. We cannot rule out that the switch induced social media editors to take more advantage of these tools than before. It is conceivable that the effectiveness of tweet promotion activities increased because of the content selection algorithm, in the sense that social media editors would be better able to target users with customized content. However, in that case, our results would still reflect an increase in user engagement due to better exposure to relevant and like-minded content.

Third, it is possible that news outlets reacted to the switch by changing the format and content of their tweets. For instance, the announcement and implementation of algorithmic curation could have raised the awareness of social media editors about the necessity to customize content for different audiences and to select stories that promise more user engagement. That is, (part of) the boost in user engagement after Twitter's switch could be explained by changes in the behavior of the newspapers rather than algorithmic selection of content by the platform. If the newspapers used the introduction of the content selection algorithm as an opportunity to adjust their tweets, we should be able to detect changes in our text-related measures. As Figure A4 shows, the daily total number of tweets slightly increased after March 16, 2016 (Panel A). However, this increase would have reduced the average user engagement per tweet for mechanical reasons rather than lead to higher per-tweet averages in likes and retweets. Importantly, Figure A4 does not indicate any other significant changes in observed tweet characteristics, nor did the hour of publication change (Figure A5). Thus, changes in newspapers' behavior are unlikely to be a dominant factor explaining our findings, at least not within our short-run window of investigation. In the long run, adjustments by news outlets are likely but this kind of investigation is beyond the scope of our data.

6.3 Differences between outlets and content

In Table 2, we investigate whether Twitter's switch had differential effects depending on the type of newspaper. For that purpose, we re-estimate Equation (2) but include triple interactions between *Twitter*, *After*, and variables that capture different newspaper characteristics. Regarding the outlets' journalistic approach (Panel A), we do not find any differences in likes between broadsheets and tabloids. However, there are some indications of effect heterogeneity in the case of retweets. As Columns (5) and (6) in Panel A show, the increase in retweets after the switch was approximately 1.4 times higher for tabloids. There are no differences when using a +/- 7-day window though (Column 4). We obtain ambiguous results when we use Wellbrock's (2011) continuous measure of newspaper quality. The coefficient on the relevant triple interaction in Panel B is not significant in most specifications. However, we find a positive association between quality and likes in case of the +/- 30-day estimation window, whereas the +/- 14-day window indicates a negative relationship between quality and retweets/shares. In terms of the outlets' scope (Panel C), all specifications consistently indicate that user engagement with tweets by national newspapers increased ca. 1.3 to 1.5 times more than user engagement with tweets by regional outlets, both in terms of likes and retweets/shares.

Figure 2 shows differences in the effect by outlets' popularity on Twitter, as measured by the number of followers on the platform before the switch. The figure summarizes the coefficients on triple interactions between *Twitter*, *After*, and quintile dummies of the number of followers, using the first quintile as the reference category.[10] These estimates indicate that user engagement increased more for more popular outlets. The differences are borderline significant for retweets. In the case of likes, outlets in the fourth and fifth quintiles experienced a significantly higher increase than outlets in the first quintile, by a factor of approximately 1.5 and 2.0, respectively (see Table A11, Panel A, for details of these regressions). We also investigate why the confidence intervals of the estimates for outlets in the fifth quintile are much larger than for the other outlets: The reason is that a small number of tweets had unusually high user engagement in the

[10] We consider the pre-treatment number of followers primarily as a moderating variable. It is also possible for the number of followers to act as a mediating variable in the effect of Twitter's switch on user engagement, especially over longer time horizons. That is, the short-run boost of the number of retweets after Twitter's switch may have increased the chances of exposing non-followers to an outlet's content. Some of these non-followers may have gradually started following the outlet, which in turn could have led to further increases in user engagement.

post-treatment period. For example, before the switch, the 10 best performing tweets in the sample on average had 189.9 retweets. After the switch, the top 10 tweets averaged 761.5 retweets, and these tweets were almost exclusively published by outlets in the fifth quintile of pre-switch outlet popularity. Hence, the introduction of the algorithm raised the chances of occasional, extraordinarily high user engagement for initially popular outlets. Tweets with a few hundred or thousand retweets can hardly be classified as "viral". However, compared to our baseline estimates (15% to 20% more user engagement), the boosts in the upper end of the distribution of likes and retweets (several hundred percent) illustrate the amplification power of algorithmic curation.

In Figure 3, we summarize estimates pertaining to the differential effects depending on topic popularity. The most popular news topics before the switch – according to the average number of likes over all tweets belonging to a topic – experienced a greater increase in user engagement than the most unpopular topics. The differences between the first and fifth quintile are borderline significant when looking at the number of likes and consistently significant for retweets. The increase in retweets was approximately 1.5 to 1.9 times larger for the initially most popular news topics than for the least popular topics, depending on the length of the estimation window (see Table A11, Panel B, for details). In general, the topic-popularity effect appears to be smaller the longer the estimation window. A possible explanation for this dissipation is that the popularity of individual news stories within topic categories (e.g., politics, sports, and business) may change relatively quickly, which our static topic model does not capture. In contrast, Twitter's algorithm might be able to detect fast-changing trends in the news cycle and update its content selection in a more dynamic manner (e.g., within a few days or hours, as compared to our +/- 30-day window), due to their access to more and better data than available to us. Hence, we are cautious to interpret the results in Figure 3 in terms of dissipation of effects over longer time horizons but mostly attribute these patterns to imperfections of our topic model.

To assess the role of quality at the level of individual news items, we evaluate interactions between *Twitter*, *After*, and various text characteristics of the tweets and posts. Table 3 summarizes the corresponding estimation results. There are no significant differences in user engagement regarding mean word length, whether the tweet/post starts with a question word, and the number of question marks. However, we obtain borderline significant results when looking at the number of words. As Columns (4) and (5) show, users were more likely to retweet shorter tweets after Twitter's switch. There are also several borderline significant coefficients regarding the number of exclamation marks, one in the case of likes (+/- 14 days) and two in the case of retweets (+/- 14 days and +/- 30 days). These point estimates indicate that user engagement increased by a factor of 1.3 after the switch if tweets included an additional exclamation mark. Importantly, we obtain multiple significant coefficients regarding the share of emotionally loaded words, both in the case of negative and positive terms and both for likes and retweets. The corresponding point estimates all indicate a positive relationship between emotional content and user engagement. Overall, Table 3 suggests that Twitter's switch to algorithmic curation increased the chances that users engage with shorter tweets and tweets that include more exclamation marks and emotionally connotated words. These text characteristics are typically associated with sensationalist reporting (Lacy and Rosenstiel, 2015; Choi, Shin, and Kang, 2021; Lischka and Garz, 2021). Thus, it can be concluded that the switch did not favor the dissemination of quality content but instead led to a propagation of tabloid-style news stories.

It is useful to consider the differences between likes and retweets when discussing the role of sensationalist content. Likes of tweets are normally not shown to others in a user's network, unless they are followers of the account whose tweet got liked. In contrast, a retweet will be featured in a user's feed, which implies that other users may be exposed to the content without having to be followers of the original publisher. From the user's perspective, the choice between liking and retweeting is often driven by psychological mechanisms. As research in social and cognitive psychology shows, people tend to share their feelings with others when they experience strong emotions.[11] Hence, users are more prone to retweet than like a tweet if the content evokes certain feelings, as it helps them to regulate their emotions. In line with this argumentation, we hardly find any differential effects of Twitter's switch on likes regarding sensationalist headlines but particularly strong effects on retweets (cp. Table 2, Panel A, and Table 3). Our interpretation is that those headlines often trigger emotional reactions among users (e.g., Grabe, Zhou, and Barnett, 2001; Uribe and Gunter, 2007). For instance, if a sensationalist headline (e.g., "Electricity bills increase by 800 percent!") makes a user feel outraged, the user might not feel particularly inclined to like the headline but instead retweet it to share the outrage with their followers.[12]

7. Conclusion

This paper analyzes the changes in user engagement with tweets published by German news outlets before and after Twitter's switch from reverse-chronological exposure to algorithmic content selection. Relying on the Facebook posts of the same outlets as a counterfactual, we use a difference-in-differences strategy to estimate the effect of the switch on users' propensity to like and retweet news content. We confirm the expectation that the introduction of the content selection algorithm boosted user engagement. According to our baseline specification, the number of likes increased by approximately 20% within a few days after the

[11] See Rimé (2009) for a review of the literature. Berger and Milkman (2012), Brady et al. (2017), and Keib, Himelboim, and Han (2018) provide empirical evidence of the role of emotions in the context of social media sharing.
[12] From the perspective of content creators, retweets are a powerful tool to amplify content. Due to their great visibility and reach, retweets often induce further retweets. They do not necessarily lead to additional likes though, given their role in emotion regulation: While retweets and likes are generally highly correlated in our data (i.e., Pearson correlation coefficient = 0.94), this relationship is much weaker when looking at the tweets by the tabloids in our sample (correlation coefficient = 0.55).

switch. The increase in the number of retweets was less pronounced, with an effect size of up to 15% after 30 days.

Importantly, our analyses indicate systematic differences across outlets and content. The evidence suggests a rich-get-richer effect, which implies that national outlets, outlets with many Twitter followers, and to some degree tabloids benefited most from the switch. Similarly, users engaged more with initially more popular news topics. We also find that user engagement increased more for shorter tweets and those including many exclamation marks and emotionally loaded words, which is compatible with the notion that sensationalist content – rather than journalistic quality – experienced a push. Overall, the findings suggest that Twitter's content selection algorithm accounts for individual preferences, but largely within mainstream content. Our results therefore resonate with pessimistic views about the relevance of the "long tail" of low-popularity, diverse media content online (Napoli, 2016).

Our study has several limitations. First, due to privacy regulations and platform policies, we do not have personal user data. Individual-level information about user preferences would be useful to investigate the political repercussions of Twitter's switch in greater detail, especially the issue of filter bubbles and belief polarization. Second, our research design does not support an analysis of the long-run effects of algorithmic curation on user engagement. It would be important to assess the performance of Twitter's algorithm over several years, as the platform likely refined the configuration of the algorithm and collected more data on user preferences. Third, our analysis is restricted to the newspaper industry in Germany. It is conceivable that Twitter's switch had different effects in countries with other political and media systems. For example, Calzada, Duch-Brown, and Gil (2022) show that algorithm updates by Google had heterogenous effects on traffic directed to the websites of news outlets across European media markets. Thus, it is advisable to be careful about generalizing any conclusions.

Our results have important implications for democratic processes though. First, the rich-get-richer effect that we find puts regional newspapers at a disadvantage. Hence, algorithmic curation could amplify the downward trend in local news that has been observed in various countries (e.g., Darr, Hitt, and Dunaway, 2018; Kekezi and Mellander, 2018; Martin and McCrain, 2019). Second, the observation that user engagement with shorter and emotionally loaded tweets increased could imply that users were more often exposed to sensationalist content with lower levels of journalistic quality. Overall, these findings imply that Twitter's switch likely had negative repercussions on citizen information, civic discourse, and accountability in politics, especially at the local level, which supports calls for cooperative responsibility between platforms, public institutions, and users (Helberger, Pierson, and Poell, 2018).

In addition, our findings have implications for the publishing industry. First, most news outlets use social media to generate advertising and subscription sales via website referrals (Hong, 2012). Mahmood and

Sismeiro (2017) estimate that a reader who likes a news story on Facebook is 25% likely to click on the story and visit the website of the publishing media outlet. According to this estimate, the newspapers in our sample may have on average experienced 5% more Twitter referrals due to the switch, if a quarter of the 20% increase in likes that we find led to page visits. However, it is doubtful whether the newspapers experienced a significant boost in revenues, as Twitter referrals normally account for less 10% of all page visits in the industry – media companies mostly use the platform to position their brand, create visibility, and influence elite discussions among politicians, intellectuals, and journalists (Cornia et al., 2018). Hence, the newspapers in our sample may have primarily benefited in non-financial terms from the switch. Second, the rich-get-richer dynamics that we find could have increased the relevance of scale effects. Media products typically have high fixed and low variable costs, which makes the industry prone to ownership concentration (e.g., Noam, 2016). Algorithmic curation could increase the likelihood of media mergers and acquisitions, because a larger number of followers raises the chances that Twitter selects content into users' feeds. Germano and Sobbrio (2020) theorize that ranking algorithms of search engines may induce few-get-richer effects, according to which websites offering unique content may attract a higher share of overall traffic. In that scenario, it could be a promising strategy for news publishers to engage in product differentiation and focus on niche audiences. However, in the context of social media – where rich-get-richer dynamics are more likely than in search – it is doubtful if niche strategies can be successful.

Acknowledgments

We thank Sebastian Axbard, Olof Åslund, Spencer Bastani, Thomas Cyron, Fabrizio Germano, Ricard Gil, Johannes Hagen, Juliane Lischka, Mattias Nordin, Jonna Rickardsson, Marco Schwarz, Shiva Shekhar, Daniel Stone, Sasha Talavera, Marshall Van Alstyne, seminar participants in Jönköping, at the 2021 Spring Meeting of Young Economists, and at the 2022 ZEW Conference on the Economics of Information and Communication Technologies for helpful suggestions and comments.

Declaration of interest statement

The authors report there are no competing interests to declare.

Data availability statement

The data and code necessary to replicate the empirical findings of this study will be made openly available on Harvard Dataverse under https://doi.org/10.7910/DVN/L5AEHV upon acceptance of the article for publication.

References

Allcott, H., Gentzkow, M., & Yu, C. (2019). Trends in the Diffusion of Misinformation on Social Media. *Research and Politics*, 6, 1–8.

Anderson, A., Maystre, L., Anderson, I., Mehrotra, R., & Lalmas, M. (2020). Algorithmic Effects on the Diversity of Consumption on Spotify. Proceedings of The Web Conference 2020, 2155–2165.

Anderson, C. (2006). *The Long Tail: Why the Future of Business is Selling Less of More*. Hyperion.

Aral, S., & Zhao, M. (2019). Social Media Sharing and Online News Consumption. Working Paper.

Backstrom, L. (2013). News Feed FYI: A Window into News Feed. Facebook for Business News, August 6, 2013. Retrieved from https://www.facebook.com/business/news/News-Feed-FYI-A-Window-Into-News-Feed

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to Ideologically Diverse News and Opinion on Facebook. *Science*, 348, 1130–1132.

Berger, J., & Milkman, K. L. (2012). What Makes Online Content Viral? *Journal of Marketing Research*, 49, 192–205.

Blank, M., & Xu, J. (2016). More Articles You Want to Spend Time Viewing. Facebook Newsroom, April 21, 2016. Retrieved from https://about.fb.com/news/2016/04/news-feed-fyi-more-articles-you-want-to-spend-time-viewing/

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022.

Bozdag, E. (2013). Bias in Algorithmic Filtering and Personalization. *Ethics and Information Technology*, 15, 209–227.

Brady, W., Wills, J., Jost, J., Tucker, J., & Van Bavel, J. (2017). Emotion Shapes the Diffusion of Moralized Content in Social Networks. *Proceedings of the National Academy of Sciences*, 114, 7313–7318.

Burbidge, J. B., Magee, L., & Robb, A. L. (1988). Alternative Transformations to Handle Extreme Values of the Dependent Variable. *Journal of the American Statistical Association*, 83, 123–127.

Cagé, J., Hervé, N., & Mayozer, B. (2020). Social Media and Newsroom Production Decisions. Working Paper.

Calzada, J., Duch-Brown, N., & Gil, R. (2022). Do Search Engines Increase Concentration in Media Markets? UB Economics Working Paper.

Cavusoglu, H., Phan, T. Q., Cavusoglu, H., & Airoldi, E. M. (2016). Assessing the Impact of Granular Privacy Controls on Content Sharing and Disclosure on Facebook. *Information Systems Research*, 27, 848–879.

Cecere, G., Jean, C., Le Guel, F., & Manant, M. (2020). STEM and Teens: An Algorithmic Bias on a Social Media. Working Paper.

Cecere, G., Corrocher, & Jean, C. (2021). Fair or Unbiased Algorithmic Decision-Making? A Review of the Literature on Digital Economics. Working Paper.

Cinelli, C. & Hazlett, C. (2020). Making Sense of Sensitivity: Extending Omitted Variable Bias. *Journal of the Royal Statistical Society B*, 82, 39–67.

Cinelli, M., Etta, G., Avalle, M., Quattrociocchi, A., Di Marco, N., Valensise, C., Galeazzi, A., & Quattrociocchi, W. (2022). Conspiracy Theories and Social Media Platforms. *Current Opinion in Psychology*, 47.

Cision (2013). "Social Journalism" Studie 2012/2013 – Deutschland. Cision Germany White Paper.

Cho, J. & Roy, S. (2004). Impact of Search Engines on Page Popularity. *Proceedings of the 13th International Conference on World Wide Web*, 2004, 20–29.

Choi, S., Shin, H., & Kang, S.-S. (2021) Predicting Audience-Rated News Quality: Using Survey, Text Mining, and Neural Network Methods. *Digital Journalism*, 9, 84–105.

Claussen, J., Peukert, C., & Sen, A. (2021). The Editor vs. the Algorithm: Targeting, Data and Externalities in Online News. Working Paper.

Cornia, A., Sehl, A., Levy, D. A. L., & Nielsen, R. K. (2018). *Private Sector News, Social Media Distribution, and Algorithm Change.* Reuters Institute.

Cowgill, B., & Tucker, C. (2020). Algorithmic Fairness and Economics. Working Paper.

Darr, J. P., Hitt, M. P., & Dunaway, J. L. (2018). Newspaper Closures Polarize Voting Behavior. *Journal of Communication*, 68, 1007–1028.

De Vries, L., Gensler, S., & Leeflang, P. S. H. (2012). Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing. *Journal of Interactive Marketing*, 26, 83–91.

DeVito, M. A. (2017). From Editors to Algorithms – A Values-Based Approach to Understanding Story Selection in the Facebook News Feed. *Digital Journalism*, 5, 753–773.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805v2.

Farkas, E. (2016). An Improved Timeline for Consumers and Brands. Twitter Blog, February 10, 2016. Retrieved from https://blog.twitter.com/en_us/a/2016/an-improved-timeline-for-consumers-and-brands.html

Flyverbom, M., Koed Madsen, A., & Rasche, A. (2017). Big Data as Governmentality in International Development: Digital Traces, Algorithms, and Altered Visibilities. *Information Society*, 33, 35–42.

Fortunato, S., Flammini, A., Menczer, F., & Vespignani, A. (2006). Topical Interests and the Mitigation of Search Engine Bias. *PNAS*, 103, 12684–12689.

Garz, M., Sörensen, J., & Stone, D. F. (2020). Partisan Selective Engagement: Evidence from Facebook. *Journal of Economic Behavior and Organization*, 177, 91–108.

Garz, M., & Szucs, F. (2022). Algorithmic Selection and Quality of Political News on Facebook. Working Paper.

Germano, F., & Sobbrio, F. (2020). Opinion Dynamics Via Search Engines (And Other Algorithmic Gatekeepers). *Journal of Public Economics*, 187.

Grabe, M. E., Zhou, S., & Barnett, B. (2001). Explicating Sensationalism in Television News: Content and the Bells and Whistles of Form. *Journal of Broadcasting & Electronic Media*, 45, 635–655.

Hagey, K., & Horwitz, J. (2021). Facebook Tried to Make Its Platform a Healthier Place – It Got Angrier Instead. Wall Street Journal, September 15, 2021.

Hanitzsch, T., Hanusch, F., Mellado, C., et al. (2011). Mapping Journalism Cultures across Nations: A Comparative Study of 18 Countries. *Journalism Studies*, 12, 273–293.

Heimbach, I., & Hinz, O. (2018). The Impact of Sharing Mechanism Design on Content Sharing in Online Social Networks. *Information Systems Research*, 29, 592–611.

Helberger, N., Pierson, J., & Poell, T. (2018). Governing Online Platforms: From Contested to Cooperative Responsibility. *Information Society*, 34, 1–14.

Holtz, D., Carterette, B., Chandar, P., Nazari, Z., Cramer, H., & Aral, S. (2020). The Engagement-Diversity Connection: Evidence from a Field Experiment on Spotify. Proceedings of the 21st ACM Conference on Economics and Computation, 75–76.

Hong, S. (2012). Online News on Twitter: Newspapers' Social Media Adoption and Their Online Readership. *Information Economics and Policy*, 24, 69–74.

Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2022). Algorithmic Amplification of Politics on Twitter. *PNAS*, 119, 1–6.

Jahr, M. (2016). Never Miss Important Tweets from People You Follow. Twitter Blog, February 10, 2016. Retrieved from https://blog.twitter.com/official/en_us/a/2016/never-miss-important-tweets-from-people-you-follow.html

Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing*. 2nd edition, Prentice Hall.

Kacholia, V., & Ji, M. (2013). Helping You Find More News to Talk About. Facebook Newsroom, December 2, 2013. Retrieved from https://about.fb.com/news/2013/12/news-feed-fyi-helping-you-find-more-news-to-talk-about/

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (2016). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives*, 5, 193–206.

Kamen, M. (2016). Twitter's Non-Chronological Timeline Is Now Opt-Out. Wired, March 17, 2016. Retrieved from https://www.wired.co.uk/article/twitter-non-chronological-timeline-how-to-opt-out

Keib, K., Himelboim, I., & Han, J.-Y. (2018). Important Tweets Matter: Predicting Retweets in the #BlackLivesMatter Talk on Twitter. *Computers in Human Behavior*, 85, 106–115.

Kekezi, O., & Mellander, C. (2018). Geography and Consumption of Local Media. *Journal of Media Economics*, 31, 96–116.

Kenter, T., & de Rijke, M. (2015). Short Text Similarity with Word Embeddings. Proceedings of the 24th ACM International Conference on Information and Knowledge Management (CIKM '15), 1411–1420.

Krug, S. (2016). Reactions Now Available Globally. Facebook Newsroom, February 24, 2016. Retrieved from https://about.fb.com/news/2016/02/reactions-now-available-globally/

Lacy, S., & Rosenstiel, T. (2015). Defining and Measuring Quality Journalism. White Paper, Rutgers School of Communication and Information.

Lambrecht, A., & Tucker, C. (2019). Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *Management Science*, 65, 2966–2981.

Lee, D., Hosanagar, K., & Nair, H. S. (2018). Advertising Content and Consumer Engagement on Social Media: Evidence from Facebook. *Management Science*, 64, 5105–5131.

Lischka, J., & Garz, M. (2021). Clickbait News and Algorithmic Curation – A Game Theory Framework of the Relation between Journalism, Users, and Platforms. *New Media & Society*, forthcoming.

Mahmood, A., & Sismeiro, C. (2017). Will They Come and Will They Stay? Online Social Networks and News Consumption on External Websites. *Journal of Interactive Marketing*, 37, 117–132.

Martin, G. J., & McCrain, J. (2019). Local News and National Politics. *American Political Science Review*, 113, 372–384.

Metallo, C., Gesuele, B., Guillamón, M.-D., & Ríos, A. M. (2020). Determinants of Public Engagement on Municipal Facebook Pages. *Information Society*, 36, 147–159.

Möller, J., Trilling, D., Helberger, N., & van Es, B. (2018). Do Not Blame it on the Algorithm: An Empirical Assessment of Multiple Recommender Systems and Their Impact on Content Diversity. *Information, Communication & Society*, 21, 959–977.

Mowshowitz, A., & Kawaguchi, A. (2002). Bias on the Web. *Communications of the ACM*, 45, 56–60.

Müller, K., & Schwarz, C. (2021). Fanning the Flames of Hate: Social Media and Hate Crime. *Journal of the European Economic Association*, 19, 2131–2167.

Myllylahti, M. (2020). Paying Attention to Attention: A Conceptual Framework for Studying News Reader Revenue Models Related to Platforms. *Digital Journalism*, 8, 567–575.

Napoli, P. M. (2016). Requiem for the Long Tail: Towards a Political Economy of Content Aggregation and Fragmentation. *International Journal of Media & Cultural Politics*, 12, 341–356.

Nikolov, D., Lalmas, M., Flammini, A., Menczer, F. (2018). Quantifying Biases in Online Information Exposure. *Journal of the Association for Information Science and Technology*, 70, 218–229.

Noam, E. (2016). *Who Owns the World's Media? Media Concentration and Ownership around the World.* Oxford University Press.

Oremus, W. (2021). Lawmakers' Latest Idea to Fix Facebook: Regulate the Algorithm. Washington Post, October 12, 2021.

Pariser, E. (2011). *The Filter Bubble: What the Internet is Hiding from You*. Penguin.

Perez, S. (2016). Twitter Says Few Users Have Opted Out of Its New, Algorithmic Timeline, March 18, 2016. Retrieved from https://techcrunch.com/2016/03/18/twitter-says-few-users-have-opted-out-of-its-new-algorithmic-timeline/

Peterson-Salahuddin, C., & Diakopoulos, N. (2020). Negotiated Autonomy: The Role of Social Media Algorithms in Editorial Decision Making. *Media and Communication*, 8, 27–38.

Peysakhovich, A., & Hendrix, K. (2016). Further Reducing Clickbait in Feed. Facebook Newsroom, August 4, 2016. Retrieved from https://about.fb.com/news/2016/08/news-feed-fyi-further-reducing-clickbait-in-feed/

Prior, M. (2003). Any Good News in Soft News? The Impact of Soft News Preference on Political Knowledge. *Political Communication*, 202, 149–171.

Read, A. (2016). Everything You Need to Know About the Latest Twitter Features (And Some Thoughts on Its Future). Buffer, March 17, 2016. Retrieved from https://buffer.com/resources/twitters-features/

Remus, R., Quasthoff, U., & Heyer, G. (2010). SentiWS – A Publicly Available German-Language Resource for Sentiment Analysis. Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010, 1168–1171.

Rimé, B. (2009). Emotion Elicits the Social Sharing of Emotion: Theory and Empirical Review. *Emotion Review*, 1, 60–85.

Rosania, P. (2015). While You Were Away. Twitter Blog, January 21, 2015. Retrieved from https://blog.twitter.com/official/en_us/a/2015/while-you-were-away-0.html

Samuelson, W., & Zeckhauser, R. (1988). Status Quo Bias in Decision Making. *Journal of Risk and Uncertainty*, 1, 7–59.

Schroeder, S. (2016). Here's How to Turn Off Twitter's New Algorithmic Timeline. Mashable, March 17, 2016. Retrieved from https://mashable.com/2016/02/10/here-s-how-to-turn-twitter-s-new-timeline-settings-on-and-off

Shmargad, Y., & Klar, S. (2020). Sorting the News: How Ranking by Popularity Polarizes Our Politics. *Political Communication*, 37, 423–446.

Sismeiro, C., & Mahmood, A. (2018). Competitive vs. Complementary Effects in Online Social Networks and News Consumption: A Natural Experiment. *Management Science*, 64, 5014–5037.

Straka, M., & Straková, J. (2017). Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, 88–99.

Stroud, N. J. (2011). *Niche News: The Politics of News Choice*. Oxford University Press.

Trilling, D., Tolochko, P., & Burscher, B. (2017). From Newsworthiness to Shareworthiness: How to Predict News Sharing Based on Article Characteristics. *Journalism and Mass Communication Quarterly*, 94, 38–60.

Uribe, R., & Gunter, B. (2007). Are ‘Sensational’ News Stories More Likely to Trigger Viewers’ Emotions than Non-Sensational News Stories? *European Journal of Communication*, 22, 207–228.

Wellbrock, C. M. (2011). Die journalistische Qualität deutscher Tageszeitungen – Ein Ranking. *Medienwirtschaft*, 8, 22–31.

Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). A Biterm Topic Model for Short Texts. *Proceedings of the 22nd International Conference on World Wide Web (WWW '13)*, 1445–1456.

Yoon, G., Li, C., Ji, Y., North, M., Hong, C., & Liu, J. (2018). Attracting Comments: Digital Engagement Metrics on Facebook and Financial Performance. *Journal of Advertising*, 47, 24–37.

Zaller, J. R. (1992). *The Nature and Origins of Mass Opinion*. Cambridge University Press.

Tables and figures

Table 1: Effects of algorithmic curation on user engagement

	Number of likes			Number of retweets/shares		
	(1)	(2)	(3)	(4)	(5)	(6)
	+/- 7 days	+/- 14 days	+/- 30 days	+/- 7 days	+/- 14 days	+/- 30 days
Twitter	0.132***	0.126***	0.129***	0.195***	0.183***	0.208***
	(0.006)	(0.004)	(0.003)	(0.015)	(0.010)	(0.009)
After	0.000	2.126	27.292	0.135	1466.416	10.069
	(0.000)	(13.176)	(115.759)	(1.456)	(14755.881)	(34.923)
Twitter × after	1.188***	1.185***	1.208***	1.104	1.071	1.151**
	(0.078)	(0.060)	(0.053)	(0.081)	(0.072)	(0.066)
Observations	25766	47931	101192	25766	47931	101192

Notes: The table shows exponentiated coefficients (i.e., incidence rate ratios) of negative binomial regressions. The column headers denote the dependent variable and estimation window. All models include a constant, day fixed effects, outlet-platform fixed effects, and an outlet-platform specific trend polynomial of order 3 (output omitted). Standard errors (in parentheses) are clustered at the outlet-platform level and refer to the exponentiated coefficients.
* p < 0.10, ** p < 0.05, *** p < 0.01

Table 2: Effects of algorithmic curation on user engagement, by newspaper characteristics

	Number of likes			Number of retweets/shares		
	(1)	(2)	(3)	(4)	(5)	(6)
	+/- 7 days	+/- 14 days	+/- 30 days	+/- 7 days	+/- 14 days	+/- 30 days
Panel A: Journalistic approach (reference: broadsheet)						
Twitter × after × tabloid	1.081	0.939	0.976	0.964	1.371***	1.404*
	(0.164)	(0.156)	(0.088)	(0.228)	(0.132)	(0.247)
Observations	25766	47931	101192	25766	47931	101192
Panel B: Wellbrock (2011) quality index (continuous)						
Twitter × after × quality	0.990	1.022	1.095**	0.972	0.889***	1.001
	(0.054)	(0.049)	(0.044)	(0.076)	(0.040)	(0.068)
Observations	19074	35702	75466	19074	35702	75466
Panel C: Scope (reference: regional)						
Twitter × after × national	1.446***	1.457***	1.355***	1.334*	1.324**	1.383***
	(0.169)	(0.136)	(0.110)	(0.200)	(0.170)	(0.149)
Observations	25766	47931	101192	25766	47931	101192

Notes: The table shows exponentiated coefficients (i.e., incidence rate ratios) of negative binomial regressions. The column headers denote the dependent variable and estimation window. All models include a constant, day fixed effects, outlet-platform fixed effects, an outlet-platform specific trend polynomial of order 3, and all constituent terms of the interactions (output omitted). Standard errors (in parentheses) are clustered at the outlet-platform level and refer to the exponentiated coefficients.
* p < 0.10, ** p < 0.05, *** p < 0.01

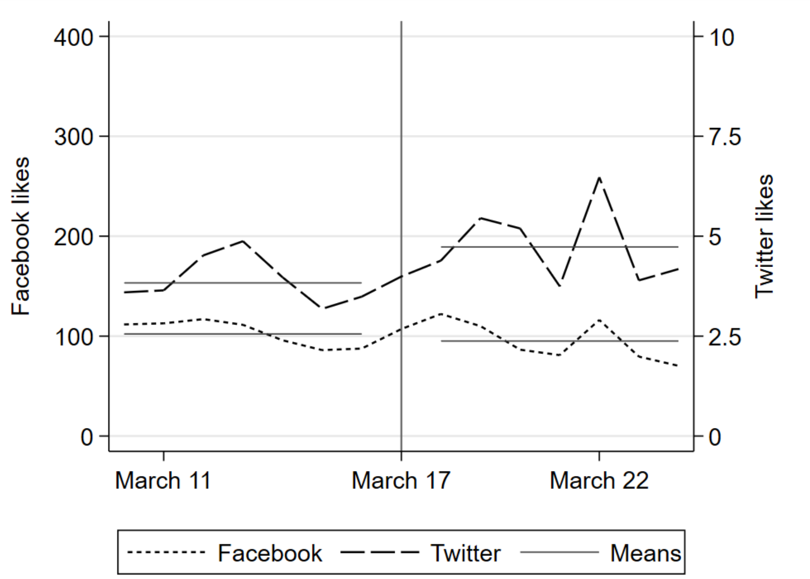
Table 3: Effects of algorithmic curation on user engagement, by text characteristics

	Number of likes			Number of retweets/shares		
	(1)	(2)	(3)	(4)	(5)	(6)
	+/- 7 days	+/- 14 days	+/- 30 days	+/- 7 days	+/- 14 days	+/- 30 days
Twitter × after × starts with question word	1.033	1.127	0.976	0.855	1.101	1.008
	(0.320)	(0.205)	(0.131)	(0.225)	(0.181)	(0.169)
Twitter × after × number of words	1.000	0.997	1.007	0.986*	0.983**	0.991
	(0.010)	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)
Twitter × after × mean word length	0.996	1.026	1.005	0.983	1.001	1.004
	(0.034)	(0.039)	(0.018)	(0.024)	(0.029)	(0.024)
Twitter × after × number question marks	1.035	1.021	0.931	1.218	1.178	0.935
	(0.112)	(0.086)	(0.067)	(0.160)	(0.120)	(0.088)
Twitter × after × number exclamation marks	1.277	1.261*	1.089	1.155	1.320*	1.382**
	(0.222)	(0.158)	(0.104)	(0.190)	(0.187)	(0.174)
Twitter × after × share negative words	3.925	2.209	1.736*	17.017***	2.896*	0.899
	(3.661)	(1.276)	(0.579)	(15.697)	(1.866)	(0.414)
Twitter × after × share positive words	6.914***	2.643***	1.142	7.800***	2.449**	0.672
	(4.538)	(0.956)	(0.462)	(5.296)	(1.021)	(0.286)
Observations	25766	47931	101192	25766	47931	101192

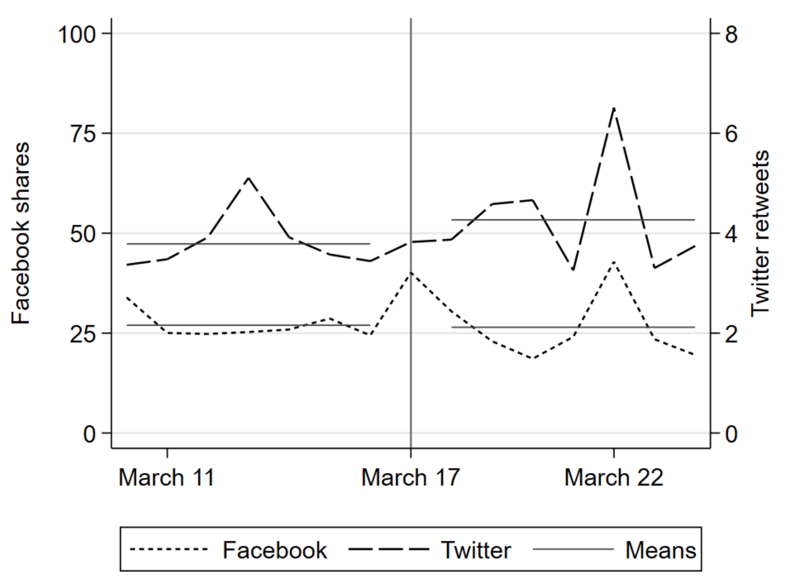
Notes: The table shows exponentiated coefficients (i.e., incidence rate ratios) of negative binomial regressions. The column headers denote the dependent variable and estimation window. All models include a constant, date fixed effects, outlet-platform fixed effects, an outlet-platform specific trend polynomial of order 3, and all constituent terms of the interactions (output omitted). Standard errors (in parentheses) are clustered at the outlet-platform level and refer to the exponentiated coefficients.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1: User engagement over time

A: Likes

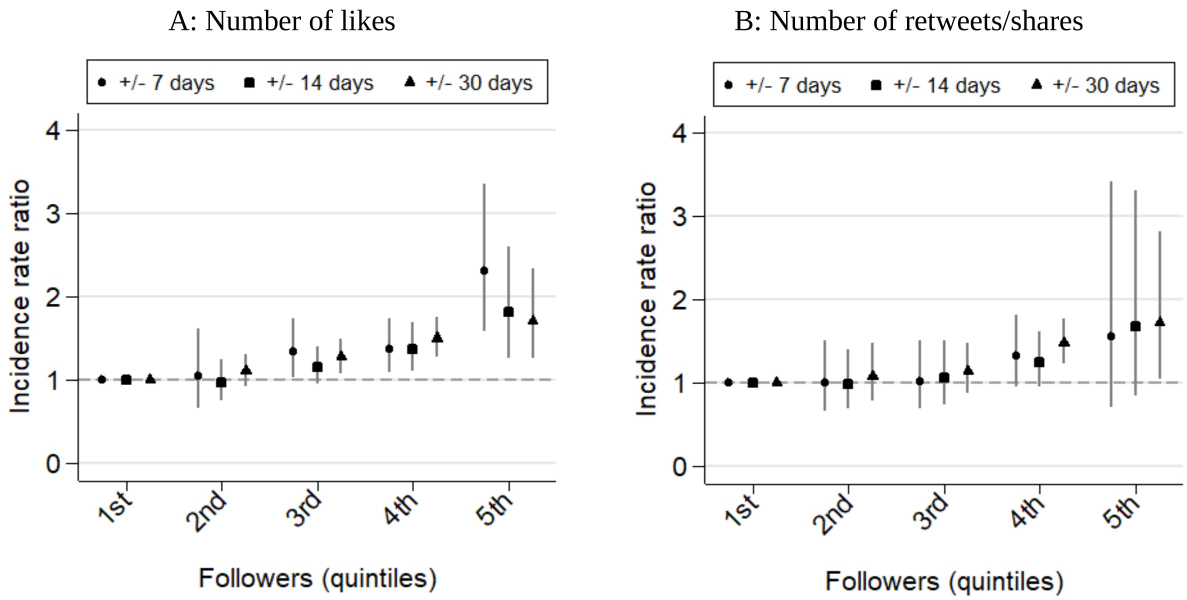


B: Retweets/shares



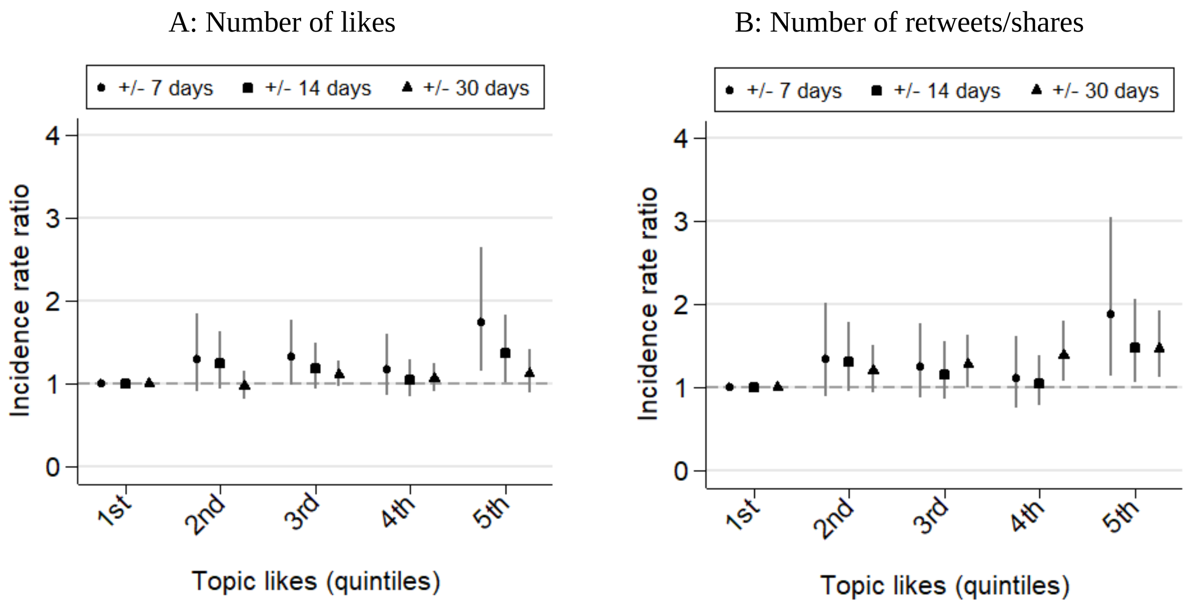
Notes: The black vertical line denotes the introduction of Twitter's content selection algorithm.

Figure 2: Effects of algorithmic curation on user engagement, by pre-switch outlet popularity



Notes: The graphs show the values of the exponentiated coefficient of regressing user engagement per tweet/post on the triple interaction between the *Twitter* dummy, the *After* dummy, and quintile dummies of the outlets' pre-switch number of Twitter followers, conditional on day and outlet-platform fixed effects and an outlet-platform specific trend polynomial of order 3 (cp. Equation 2). The first quintile is the reference category. The grey solid spikes denote the 95% confidence interval, based on standard errors clustered at the outlet-platform-level.

Figure 3: Effects of algorithmic curation on user engagement, by pre-switch topic popularity



Notes: The graphs show the values of the exponentiated coefficient of regressing user engagement per tweet/post on the triple interaction between the *Twitter* dummy, the *After* dummy, and quintile dummies of the topics' average number of likes before the switch, conditional on day and outlet-platform fixed effects and an outlet-platform specific trend polynomial of order 3 (cp. Equation 2). The first quintile is the reference category. The grey solid spikes denote the 95% confidence interval, based on standard errors clustered at the outlet-platform-level.

Online Supplementary Appendix

Table A1: List of outlets

Outlet	Twitter handle	Facebook domain
Abendblatt Hamburg	abendblatt	abendblatt
Berliner Morgenpost	morgenpost	morgenpost
Berliner Zeitung	berlinerzeitung	berlinerzeitung
B.Z. Berlin	bzberlin	B.Z.Berlin
Express	express24	EXPRESS.Koeln
Frankfurter Allgemeine	faznet	faz
Focus	focusonline	focus.de
Frankfurter Rundschau	fr	FrankfurterRundschau
Freie Presse	freie_presse	freiepresse
General-Anzeiger	gabonn	gaonline
Handelsblatt	handelsblatt	handelsblatt
Kölner Stadt-Anzeiger	ksta_news	ksta.fb
Lausitzer Rundschau	lr_online	lausitzerrundschau
Leipziger Volkszeitung	LVZ	lvzonline
Main-Post	mainpost	mainpost
Märkische Allgemeine	maz_online	MAZonline
Mitteldeutsche Zeitung	mzwebde	mzwebde
MOPO	mopo	hamburgermorgenpost
Neue Westfälische	nwnews	NeueWestfaelische
Nürnberger Zeitung	nordbayern	nordbayern.de
Nordwest-Zeitung	nwzonline	nwzonline
Neue Osnabrücker	noz_de	neueoz
Ostthüringer Zeitung	OTZonline	otz.de
Passauer Neue Presse	pnp	pnp.de
Rheinische Post	rponline	rponline
Schwäbische Zeitung	Schwaebische	schwaebische.de
Der Spiegel	DerSPIEGEL	DerSpiegel
Süddeutsche Zeitung	SZ	ihre.sz
Südkurier	Suedkurier_News	Suedkurier.News
Südwest Presse	SWPde	swp.de
Der Tagesspiegel	Tagesspiegel	Tagesspiegel
Die Tageszeitung	tazgezwitscher	taz.kommune
Thüringische LZ	TLZnews	tlz.de
Thüringer Allgemeine	TAOnline	thueringerallgemeine
Die Welt	welt	welt
Die Welt Kompakt	weltkompakt	weltkompakt
Die Zeit	DIEZEIT	diezeit

Table A2: Summary statistics of outlet-level variables

	Mean	SD	Min.	Max.	Obs.
Type					
-tabloid (binary)	0.08	0.28	0.00	1.00	37
-broadsheet (binary)	0.92	0.28	0.00	1.00	37
Wellbrock (2011) quality index	6.86	0.94	4.67	8.38	26
Scope					
-national (binary)	0.27	0.45	0.00	1.00	37
-regional (binary)	0.73	0.45	0.00	1.00	37
Number of pre-switch Twitter followers	123872.49	199444.43	1306.00	806916.00	37

Table A3: Summary statistics of tweet/post-level variables

	Twitter		Facebook	
	Mean	SD	Mean	SD
User engagement				
-number of likes	4.404	17.743	106.661	997.666
-number of retweets/shares	4.112	23.709	30.910	334.047
Message characteristics				
-starts with question word (binary)	0.033	0.178	0.047	0.212
-number of words	11.826	3.627	16.079	13.086
-mean word length	11.610	5.146	6.930	1.368
-number of question marks	0.173	0.404	0.175	0.438
-number of exclamation marks	0.071	0.283	0.180	0.463
-share of negative words	0.042	0.065	0.036	0.069
-share of positive words	0.039	0.061	0.052	0.081
Observations	56621		44571	

Notes: The observations include all tweets and posts within a window of +/-30 days around Twitter's switch to algorithmic curation.

Table A4: Characteristics of news topics

Label	Share in corpus (%)	Avg. Twitter likes before switch	Top terms
Sports	8.83	2.43	game, soccer, celebrate, close, last, second, victory, bundesliga, coach, fan, win, lose, week, live ticker, player, series, season, end, minute, goal
Refugee crisis	8.35	2.69	refugee, political party, demand, choice, politically, refugee crisis, state election, refugee policy, border, stay, merkel, government, summit, strong, dispute, demonstrate, citizen, plan, survey, important
Panorama	8.32	2.87	human, child, all, world, man, life, young, money, actually, knowledge, own, old, family, help, correct, parents, house, city, anxiety, fast
Call to actions	8.24	3.10	picture, find, question, important, video, see, photo, far, thanks, read, note, reply, say, interview, write, tell, speak, colleague, sad, network
Weather	6.93	3.71	well, tomorrow, day, night, beautiful, weekend, weather, sun, stay, photo, enjoy, long, today, early, great, spring, time, goodbye, yesterday, bad
Politics	6.84	3.08	turkish, demand, president, boss, politician, criticism, care, government, federal government, merkel, explain, threaten, dispute, comment, politics, decision, police, foreign minister, prime minister, chancellor
Traffic news	6.18	2.00	fire, train, police, road, automobile, full, flat, build, free, night, drive, direction, burn, completely, city, airport, place, traffic, fire department, building, accident
Accidents	5.76	2.42	heavy, hurt, accident, car, die, child, driver, aged, dead, kill, rescue, fatal, girl, police, animal, boy, injured, mother, meter, hospital
Business	5.62	1.96	million, money, high, percent, public, numbers, get, billion, employee, receive, problem, costs, companies, future, expensive, customer, bank, job, private, service
Crime investigations	5.37	2.23	police, search, perpetrator, presumed, officer, arrest, witness, unknown, law, drug, victim, threaten, brutally, pull, beat, suspects, knife, raid, assault, solve
Tech news	4.87	2.43	video, put, get, see, page, call, start, online, work, current, study, knowledge, user, book, internet, product, technology, path, manufacturer, map
Economy	4.41	2.80	large, climb, number, clear, year, strong, price, percent, lie, stay, expect, sink, calculate, reason, grow, scarce, federal state, fast, region, reach
Foreign news	4.14	2.09	attack, explosion, belgian, police, border, apparently, third, german, grasp, identify, terrorist, report, human, macedonian, turkish, greek, medium, confirm, authority, capital
Crime committed by refugees	3.68	4.08	refugee, sexually, child, court, violence, police, legal action, reproach, broadcasting fee, conviction, numbers, throw, doctor, russian, dispute, family, judge, social, fight, assault, lawyer
Court news	3.48	2.09	detention, court, condemn, prosecutor, process, past, determine, murder, detection, accusation, suspicion, jail, judgment, defendant, dismiss, begin, district court, load, procedure, prison sentence
Corporate crime	3.25	2.53	warn, struggle, business, error, testing, researcher, mailbox companies, controversial, failure, conflict, danger, company, panama papers, terror, knowledge, establish, scandal, tax havens, politics, allegedly
Outlets' own matters	2.71	2.50	sports, link, newspaper, taz, tagesspiegel, green, beer, red, magazine, süddeutsche, fohlenfutter, water, interview, photo, comment, write, take, handelsblatt, introduce, lead
Highlights of the print edition	2.21	2.39	zeit, discover, view, plane, find, political part, edition, short department, print, crash, economic section, object, year, page, chance, carry, article, last, leave, week

Notes: The topic shares do not sum up to 100% because a fraction of tweets/posts have an empty message text, in which case it is not possible to assign any topic. The top terms are English translations, after excluding overly common terms.

Table A5: Effects of algorithmic curation on user engagement, using samples of matched and unmatched tweets/posts

	Number of likes			Number of retweets/shares		
	(1)	(2)	(3)	(4)	(5)	(6)
	+/- 7 days	+/- 14 days	+/- 30 days	+/- 7 days	+/- 14 days	+/- 30 days
Panel A: Tweets/posts with counterpart on other platform						
Twitter × after	0.956	1.057	1.232***	1.257	1.285***	1.293***
	(0.145)	(0.109)	(0.084)	(0.194)	(0.104)	(0.083)
Partial R^2 of treatment	0.001	0.004	0.117	0.030	0.118	0.185
Robustness value	0.034	0.061	0.304	0.162	0.305	0.376
Observations	1852	3591	7417	1852	3591	7417
Panel B: Tweets/posts without counterpart on other platform						
Twitter × after	1.187**	1.254***	1.215***	1.187**	1.102	1.123*
	(0.095)	(0.076)	(0.063)	(0.087)	(0.070)	(0.069)
Partial R^2 of treatment	0.061	0.162	0.165	0.071	0.031	0.047
Robustness value	0.225	0.354	0.357	0.240	0.164	0.199
Observations	10990	21115	44972	10990	21115	44972
Panel C: All tweets/posts						
Twitter × after	1.188***	1.185***	1.208***	1.104	1.071	1.151**
	(0.078)	(0.060)	(0.053)	(0.081)	(0.072)	(0.066)
Partial R^2 of treatment	0.087	0.135	0.206	0.025	0.014	0.076
Robustness value	0.265	0.325	0.396	0.147	0.113	0.248
Observations	25766	47931	101192	25766	47931	101192

Notes: The table shows exponentiated coefficients (i.e., incidence rate ratios) of negative binomial regressions. The column headers denote the dependent variable and estimation window. The sample used in Panel A includes posts and tweets with a cosine similarity larger than 0.8, whereas Panel B includes posts and tweets with a cosine similarity below 0.2. For comparison, Panel C shows the same results as from the baseline model in Table 1. All models include a constant, day fixed effects, and outlet-platform fixed effects (output omitted). *Partial R^2 of treatment* indicates how much of the variation in the engagement metrics is explained by Twitter's switch, whereas *Robustness value* indicates how strongly unobserved confounders would have to be associated with both user engagement and the treatment – measured in terms of partial R^2 – to drive the treatment effect to zero (see Cinelli and Hazlett, 2020). Standard errors (in parentheses) are clustered at the outlet-platform level and refer to the exponentiated coefficients.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Effects of algorithmic curation on user engagement (OLS estimates I)

	Log(number of likes + 1)			Log(number of retweets/shares + 1)		
	(1)	(2)	(3)	(4)	(5)	(6)
	+/- 7 days	+/- 14 days	+/- 30 days	+/- 7 days	+/- 14 days	+/- 30 days
Twitter	-1.944***	-1.995***	-2.050***	-1.114***	-1.179***	-1.161***
	(0.014)	(0.013)	(0.014)	(0.015)	(0.017)	(0.014)
After	-5.150	0.185	1.745*	0.802	3.244	0.396
	(3.417)	(1.989)	(1.029)	(2.870)	(4.033)	(0.964)
Twitter × after	0.059**	0.047**	0.079***	0.074**	0.025	0.017
	(0.028)	(0.023)	(0.025)	(0.031)	(0.030)	(0.026)
Adjusted R^2	0.705	0.702	0.700	0.403	0.398	0.397
Observations	25766	47931	101192	25766	47931	101192

Notes: The column headers denote the dependent variable and estimation window. All models include a constant, day fixed effects, outlet-platform fixed effects, and an outlet-platform specific trend polynomial of order 3 (output omitted). Standard errors (in parentheses) are clustered at the outlet-platform level.
* p < 0.10, ** p < 0.05, *** p < 0.01

Table A7: Effects of algorithmic curation on user engagement (OLS estimates II)

	Inverse hyperbolic sine of likes			Inverse hyperbolic sine of retweets/shares		
	(1)	(2)	(3)	(4)	(5)	(6)
	+/- 7 days	+/- 14 days	+/- 30 days	+/- 7 days	+/- 14 days	+/- 30 days
Twitter	-2.134***	-2.199***	-2.272***	-1.308***	-1.395***	-1.373***
	(0.016)	(0.015)	(0.016)	(0.018)	(0.020)	(0.016)
After	-6.554	-0.205	1.968	0.539	3.747	0.379
	(3.945)	(2.204)	(1.208)	(3.493)	(4.909)	(1.264)
Twitter × after	0.067**	0.054**	0.084***	0.086**	0.025	0.011
	(0.031)	(0.026)	(0.028)	(0.036)	(0.036)	(0.030)
Adjusted R^2	0.720	0.717	0.714	0.412	0.406	0.406
Observations	25766	47931	101192	25766	47931	101192

Notes: The column headers denote the dependent variable and estimation window. All models include a constant, day fixed effects, outlet-platform fixed effects, and an outlet-platform specific trend polynomial of order 3 (output omitted). Standard errors (in parentheses) are clustered at the outlet-platform level.
* p < 0.10, ** p < 0.05, *** p < 0.01

Table A8: Effects of algorithmic curation on likes (placebo regressions)

	(1)	(2)	(3)	(4)
	Fictional treatment date:			
	12 months before switch	3 months before switch	3 months after switch	12 months after switch
Twitter	0.037***	0.082***	0.117***	0.000***
	(0.002)	(0.015)	(0.005)	(0.000)
After	17.060**	0.745	0.507	61.106**
	(22.790)	(1.927)	(0.288)	(121.087)
Twitter × after	0.983	0.936	0.984	1.018
	(0.065)	(0.064)	(0.066)	(0.064)
Observations	18789	22749	26518	31474

Notes: The table shows exponentiated coefficients (i.e., incidence rate ratios) of negative binomial regressions. Dependent variable: number of likes. The column headers denote fictional treatment dates. All estimates refer to the +/- 7-day windows around these dates. All models include a constant, day fixed effects, outlet-platform fixed effects, and an outlet-platform specific trend polynomial of order 3 (output omitted). Standard errors (in parentheses) are clustered at the outlet-platform level and refer to the exponentiated coefficients.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A9: Effects of algorithmic curation on retweets/shares (placebo regressions)

	(1)	(2)	(3)	(4)
	Fictional treatment date:			
	12 months before switch	3 months before switch	3 months after switch	12 months after switch
Twitter	0.077***	0.171***	0.148***	0.000***
	(0.005)	(0.026)	(0.009)	(0.000)
After	10.260	2.681	1.140	33.259*
	(18.928)	(5.529)	(1.132)	(65.876)
Twitter × after	1.075	0.946	0.922	0.867
	(0.080)	(0.070)	(0.061)	(0.078)
Observations	18789	22749	26518	31474

Notes: The table shows exponentiated coefficients (i.e., incidence rate ratios) of negative binomial regressions. Dependent variable: number of retweets/shares. The column headers denote fictional treatment dates. All estimates refer to the +/- 7-day windows around these dates. All models include a constant, day fixed effects, outlet-platform fixed effects, and an outlet-platform specific trend polynomial of order 3 (output omitted). Standard errors (in parentheses) are clustered at the outlet-platform level and refer to the exponentiated coefficients.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A10: Effects of algorithmic curation on daily total user engagement

	Number of likes			Number of retweets/shares		
	(1)	(2)	(3)	(4)	(5)	(6)
	+/- 7 days	+/- 14 days	+/- 30 days	+/- 7 days	+/- 14 days	+/- 30 days
Twitter	0.195***	0.194***	0.198***	0.289***	0.288***	0.295***
	(0.041)	(0.030)	(0.025)	(0.049)	(0.034)	(0.026)
After	0.000	26.680	0.003	0.034	146.312	0.015***
	(0.001)	(159.909)	(0.012)	(0.215)	(1032.437)	(0.019)
Twitter × after	1.169**	1.140**	1.152***	1.128*	1.129*	1.161***
	(0.089)	(0.066)	(0.046)	(0.078)	(0.078)	(0.053)
Observations	1072	2049	4263	1072	2049	4263

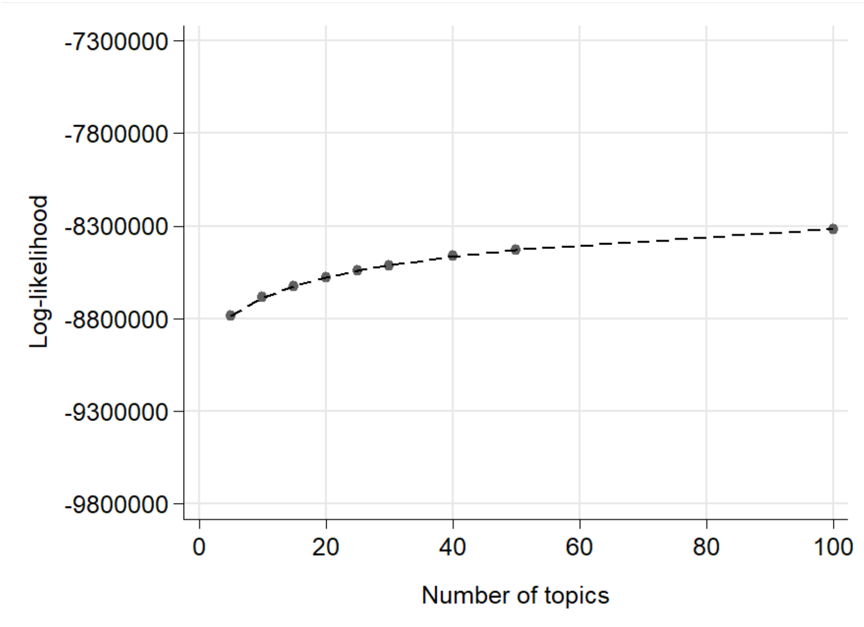
Notes: The table shows exponentiated coefficients (i.e., incidence rate ratios) of negative binomial regressions, using data at the outlet-platform-day level. The dependent variable is the sum of likes or sum of retweets/shares per outlet over all tweets or posts on a given day. All models include a constant, day fixed effects, outlet-platform fixed effects, and an outlet-platform specific trend polynomial of order 3 (output omitted). Standard errors (in parentheses) are clustered at the outlet-platform level and refer to the exponentiated coefficients.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A11: Effects of algorithmic curation on user engagement, by outlet and content popularity

	Number of likes			Number of retweets/shares		
	(1)	(2)	(3)	(4)	(5)	(6)
	+/- 7 days	+/- 14 days	+/- 30 days	+/- 7 days	+/- 14 days	+/- 30 days
Panel A: Number of followers (reference category: 1st quintile)						
Twitter × after × 2nd quintile	1.044	0.972	1.108	1.002	0.985	1.083
	(0.236)	(0.125)	(0.098)	(0.209)	(0.179)	(0.171)
Twitter × after × 3rd quintile	1.346**	1.163	1.278***	1.027	1.060	1.142
	(0.180)	(0.110)	(0.105)	(0.201)	(0.192)	(0.152)
Twitter × after × 4th quintile	1.380***	1.370***	1.504***	1.321*	1.251*	1.481***
	(0.165)	(0.148)	(0.121)	(0.218)	(0.167)	(0.133)
Twitter × after × 5th quintile	2.304***	1.815***	1.715***	1.554	1.683	1.727**
	(0.442)	(0.337)	(0.271)	(0.625)	(0.583)	(0.434)
Observations	25766	47931	101192	25766	47931	101192
Panel B: Initial topic popularity (reference category: 1st quintile)						
Twitter × after × 2nd quintile	1.296	1.245	0.976	1.348	1.307*	1.201
	(0.235)	(0.173)	(0.085)	(0.276)	(0.209)	(0.144)
Twitter × after × 3rd quintile	1.327*	1.188	1.116	1.252	1.162	1.279**
	(0.199)	(0.138)	(0.079)	(0.222)	(0.172)	(0.160)
Twitter × after × 4th quintile	1.181	1.046	1.067	1.113	1.045	1.392**
	(0.185)	(0.114)	(0.085)	(0.213)	(0.151)	(0.183)
Twitter × after × 5th quintile	1.749***	1.367**	1.126	1.874**	1.483**	1.472***
	(0.368)	(0.207)	(0.132)	(0.467)	(0.249)	(0.200)
Observations	25766	47931	101192	25766	47931	101192

Notes: The table shows exponentiated coefficients (i.e., incidence rate ratios) of negative binomial regressions. The column headers denote the dependent variable and estimation window. All models include a constant, day fixed effects, outlet-platform fixed effects, an outlet-platform specific trend polynomial of order 3, and all constituent terms of the interactions (output omitted). Standard errors (in parentheses) are clustered at the outlet-platform level and refer to the exponentiated coefficients.
* p < 0.10, ** p < 0.05, *** p < 0.01

Figure A1: Fit of biterm topic models with different numbers of topics



Notes: Lower absolute values reflect a better model fit. The log-likelihood refers to the sum of log-likelihoods of all biterms in a model.

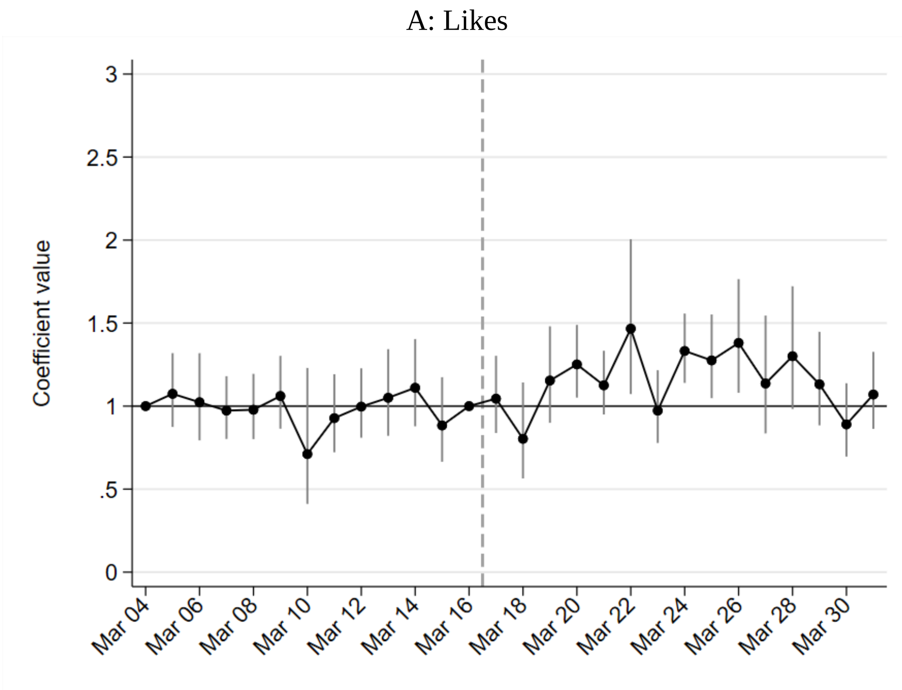
Figure A2: Examples of (un-)matched tweets and posts

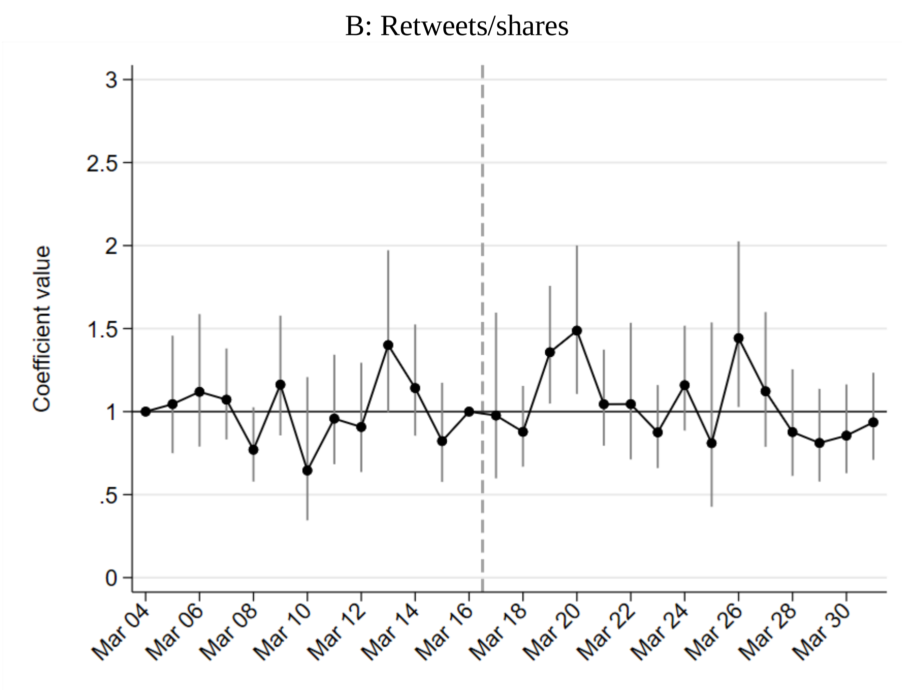
A: matched (cosine similarity = 0.861)

B: unmatched (cosine similarity = 0.125)

Notes: We consider a tweet or post to have a cross-platform twin if their maximum cosine similarity with the items on the other platform on the same day by the same newspaper is larger than 0.8, whereas we consider tweets/posts to have no cross-platform twin when the maximum cosine similarity is smaller than 0.2.

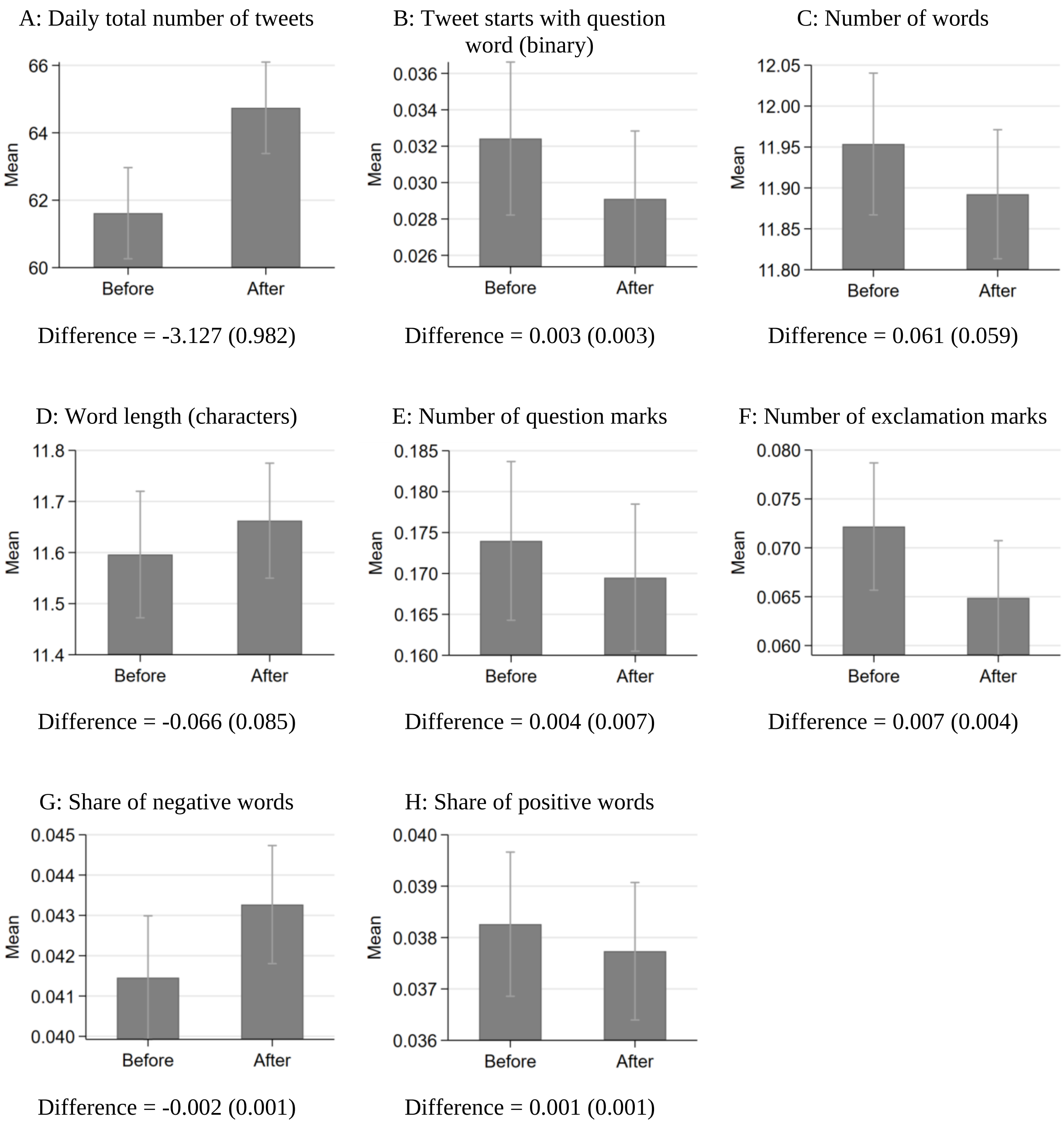
Figure A3: Estimated coefficients on interactions between the Twitter indicator and time dummies





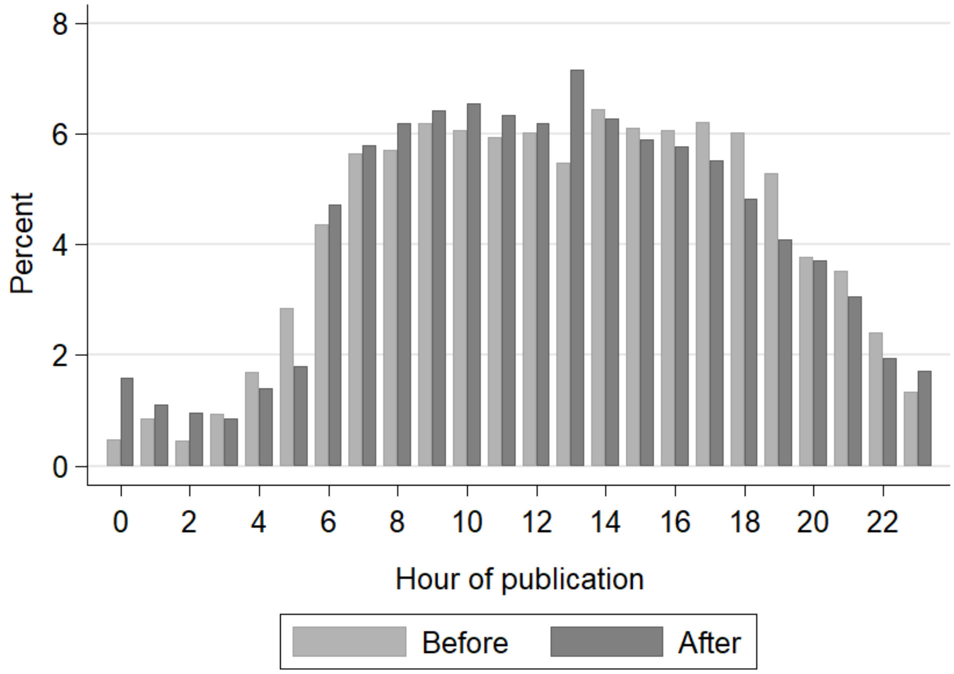
Notes: The graph shows the exponentiated coefficients from negative binomial regressions of the number of likes and number of retweets/shares, respectively, on interactions between the Twitter dummy and time dummies, conditional on outlet-platform fixed effects. The grey dashed line marks the start of algorithmic curation on March 16, 2016. The grey solid spikes denote the 95% confidence interval, based on standard errors clustered at the outlet-platform-level.

Figure A4: Characteristics of tweets before and after the introduction of algorithmic curation



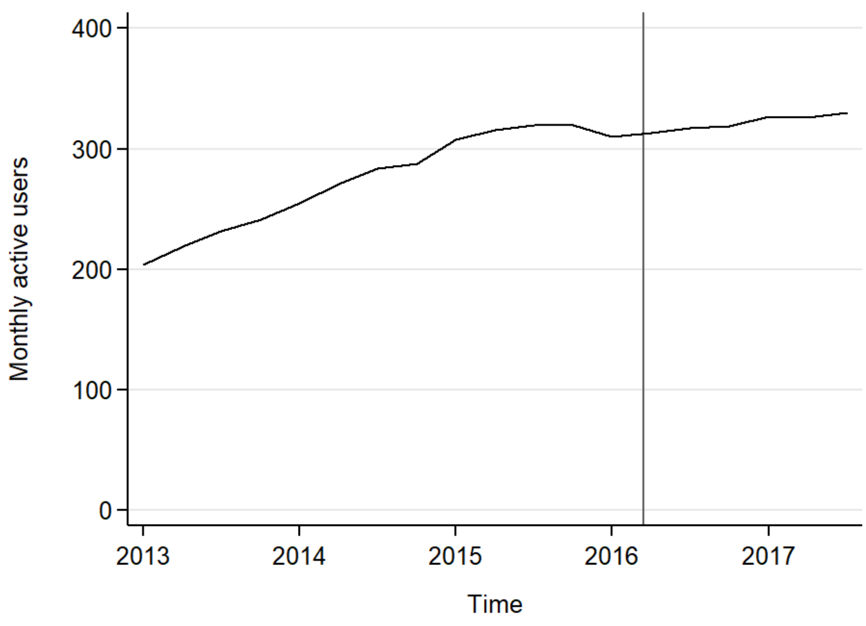
Notes: Observations include all tweets within a window of +/-7 days around the switch. The error bars represent the 95% confidence interval. The values in parentheses denote the standard error of the before-after difference.

Figure A5: Time of publication of tweets, before and after the introduction of algorithmic curation



Notes: Observations include all tweets within a window of +/-7 days around the switch.

Figure A6: Global development of the number of Twitter users, 2013 – 2017



Notes: Data based on Twitter's annual reports to stakeholders. The black vertical line denotes the introduction of Twitter's content selection algorithm.