Do the Media Bow to Foreign Economic Powers? Evidence from a News Website Crackdown*

HENG CHEN
University of Hong Kong
LI HAN
Hong Kong University of Science and Technology

February 7, 2022

Abstract. Do news media compromise their reporting to maintain market access to China? Exploiting a large-scale media crackdown in May 2019, in which multiple influential UK- and US-based news sites were blocked, we find that media outlets adopted a more negative tone in news reporting on China after being blocked, compared to those with no access change. The negative effect is present only in news on politically sensitive topics and is absent in news on economic topics and opinion articles. The crackdown also led to a higher frequency of reporting on sensitive topics such as human rights; however, again, reporting on economic topics remained unaffected. Further evidence suggests that these findings are aligned with the interpretation that the media censored themselves less after losing access, but not with the conjecture that they retaliated against China or responded to the changed readership.

Keywords: Self-censorship; Market Access; Crackdown; Word Embedding; Topic Modeling; Computational Linguistics

JEL Classification. L82 L88 F69 P00

*Heng CHEN: The Faculty of Business and Economics, The University of Hong Kong; Email: hengchen@hku.hk. Li HAN: Division of Social Sciences, The Hong Kong University of Science and Technology; Email: lihan@ust.hk. We thank Haohan Chen, Ting Chen, Jean Hong, Chicheng Ma, Ruixue Jia, Wing Suen, Yikai Wang, and Wen Wang for useful comments and discussions. We are grateful for the excellent research assistance provided by Ziwei Wang, Xinyu Wei, and Guanting Yi. This research project is partly funded by the General Research Fund of the Research Grants Council of Hong Kong (Project No. HKU 17500421 and Project No. HKUST 16503819) and by the Seed Fund for Basic Research of Hong Kong University (Project No. 201811159002).

1. Introduction

In an era of unprecedented economic integration of democratic and nondemocratic countries, a new dilemma has arisen for media outlets that operate globally. On the one hand, the norm of objectivity in journalism and concerns over reputation require that the media truthfully inform citizens in democracies about international issues, and this function is essential for foreign policymaking. On the other hand, the recent economic rise of authoritarian countries has empowered their governments to influence news reporting in their own favor. Do the media succumb to such influence, especially when commercial interests are at stake?

This concern is not unwarranted. For example, according to NPR, Bloomberg News "killed an investigation into the wealth of Communist Party elites in China, fearful of repercussions by the Chinese government" in 2013.[1] Bloomberg's editor-in-chief justified this editorial decision in a private (but taped and eventually leaked) conference call with the outlet's China-based investigative team:

> "It is for sure going to, you know, invite the Communist Party to, you know, completely shut us down and kick us out of the country. So I just don't see that as a story that is justified."

The editor went on in the same conference call to suggest a compromise strategy to deal with the dilemma at hand:

> "There's a way to use the information you have in such a way that enables us to report, but not kill ourselves in the process and wipe out everything we've tried to build there."

The editor's comment accentuated one's apprehension of losing access to the market, a concern likely shared by other media outlets. The value of access to large markets in authoritarian countries can hardly be ignored. It is clearly embodied in the decade-long struggle of the New York Times to gain and regain access to China (elaborated in section 2.3), as well as Facebook founder Zuckerberg's undisguised effort to charm Chinese censors into permitting the company's entry.[2] Consequently, access has become a source of leverage that authoritarian governments can wield over foreign media. In this context, China is far from an exception. Vietnam, another

[1]See "Bloomberg News Killed Investigation, Fired Reporter, Then, Sought To Silence His Wife." April 14, 2020, NPR.

[2]It is in the interest of the media to cultivate audiences and strengthen their brands in foreign markets, in particular, in those of populous and fast-growing countries. See "The New York Times vs. the 'Great Firewall' of China"(March 31, 2017, The New York Times). On Mark Zuckerberg's effort, see "Facebook Gains Status in China, at Least for a Moment"(July 24, 2018, The New York Times).

fast-growing authoritarian country, has blatantly coerced Facebook and Google into censorship with the threat of shutting them out of the country.[3] The news business is unlikely to be the only victim under the authoritarian governments' influence either: reportedly, "[t]o preserve their access to the country, institutions from Apple and Hollywood to the International Olympic Committee (ioc) remain silent on all matters sensitive to the Communist Party."[4]

This paper investigates whether news media systematically compromise their reporting to make their way into economically important authoritarian markets. We address this issue in the context of China. Specifically, we study *whether* news organizations based in democratic countries tone down or suppress negative information about China to appease censors and maintain market access and examine the *areas in which* they compromise.

It is challenging to isolate the effect of market access because gaining (or losing) access is likely to be endogenous to the content published by news outlets. To deal with this challenge, we exploit a large-scale "rectification" campaign launched by the Chinese government in the middle of 2019, in which major foreign news outlets, and a large number of Chinese social media and Chinese financial news websites were blocked. The purpose of the crackdown was to control information on the causes and consequences of the unexpected breakdown of trade negotiations between the US and China. In this particular campaign, the blockage of news websites was based on their influence in China rather than the content of their reports; this feature allows us to use the difference-in-differences model to identify the impact of losing access on media outlets' reporting strategy. Specifically, we compare the change in the tone and frequency of reporting by blocked outlets before and after the campaign with that of outlets with no access change in the same period, and explore whether those changes differ across topics.

We focus on news and opinion articles on China published by major US and UK news outlets in the period from January 2018 to May 2020. Most news media have an opinion section publishing articles with subjective views, including opinions, letters from readers, op-eds, and contributions from columnists, and its editorial operation is independent from that of news sections. Therefore, we examine news and opinion articles separately.

Six major outlets that had salient presence in China and publish English content were blocked in June 2019 in the aforementioned campaign. We label them the treat-

[3]See "Facebook and YouTube accused of complicity in Vietnam repression (December 1, 2020, The Guardian)" and "Vietnam threatens to shut down Facebook over censorship requests" (November 20, 2020, Reuters).

[4]See "Cold warrior: why Eileen Gu ditched Team USA to ski for China (February 3rd, 2022, The Economist)"

ment group. To construct the control group, we put together English-language news outlets with comparable circulation volumes and influence but no change in access to China. These criteria leave us 13 outlets, including those always blocked and those never blocked during our data period.

It is difficult to measure the reporting strategy systematically across diverse content, therefore, we focus on the news coverage frequency and news tone, the two main characteristics (i.e., extensive and intensive margins). The advantage of studying the news tone is that it can be compared across time, outlets, topics, and articles. The change in tone can at least serve as a conservative measure of the media's adjustment in their handling of China-related news. Using the word embedding method, we compute word-level tone scores. Then, we aggregate such scores to construct article-level tone scores as our main measure of news tone.

Our analysis shows that the treated media indeed changed their China reporting strategy. Relative to China-related news articles published by outlets in the control group, such articles published by the treated outlets assumed a more negative tone after the 2019 blockage. The negative impact is statistically significant even if standard errors are estimated using the cluster-adjusted wild bootstrapping and randomization inference approaches, which relieves our concern about the potential bias caused by the relatively small number of sample outlets. Interestingly, no similar pattern is observed for opinion articles.

Could the result be driven by the response of the never-blocked media to the crackdown? We rule out this concern by showing that our result is robust to removing the never-blocked outlets from the sample and that their tone did not change differently from that of the always-blocked media after the crackdown.

One threat to our identification strategy is that the blockage was endogenous to the media content or the preexisting trend of the latter. To address this concern, we first show that our result is robust to excluding articles related to the actual or suspected triggers of the crackdown — the unexpected fallout of Sino-US trade negotiations or the 30th anniversary of the Tiananmen Incident. Second, using an event study model, we show that there was no difference in pre-trends between the treatment and control groups and that the change in the tone coincided with the crackdown. These findings reassure us that the crackdown was not endogenous to the content.

Concerns may remain that the results are confounded by time-varying outlet-specific factors. In particular, treated media could be more responsive to newsworthy events in relation to authoritarian politics after the crackdown. To address this concern, we first show that the result is robust to excluding news articles related to prominent issues such as the Hong Kong protests and COVID-19. Next, we consider Russia- and Iran-related articles as an additional comparison group and use a difference-in-differences-in-differences (DDD) model to demonstrate that there were no outlet-specific changes toward authoritarian regimes. These checks corroborate the idea that the blockage led to the changes in the media's reporting strategy.

Was the negative effect present on all news topics related to China or on only a subset of them before the crackdown? Estimating a Latent Dirichlet Allocation (LDA) topic model with our news corpus, we discover thirteen interpretable news topics. We find that the negative effect of blockage on news tone consistently arises for reporting on politically sensitive topics such as human rights but not for politically non-sensitive topics such as economic growth.

Did the heterogeneity in changes for sensitive and non-sensitive topics also occur in the case of coverage frequency? The treated media are indeed found to publish more articles on sensitive topics (i.e., human rights, the Sino-US relationship and Huawei-related high-tech security issues) after the crackdown, compared to the control group. In contrast, no similar pattern arises for non-sensitive economic topics. Considering all topics together, the treated media outlets produced more news articles about China after the crackdown, but the difference is not statistically significant.

The findings above are aligned with the interpretation that the crackdown removed a constraint on media outlets and made them less concerned about upsetting Chinese censors than when they strived to maintain access. In other words, news outlets, before being blocked, may have intentionally softened the tone toward China in their news reporting or even chose to report less often on sensitive topics. However, such intentional effort may not have been applied to economic news, as they are not sensitive, or opinion sections, as the media claim no responsibility for perspectives expressed in opinion articles.

Alternatively, the tone and frequency changes may be interpreted as the media retaliating against the Chinese government or responding to changes in the composition of readers. Albeit plausible, none of the alternative explanations hold water, given a closer look at the responses of the media. Media outlets that had more commercial interests or influence in China and hence suffered more from losing the Chinese market or readers actually responded more mildly — rather than more vehemently — to the blockage than did other outlets; this conflicts with the retaliation interpretation. Once we incorporate proxies for readership in our analysis, we find that the tone change resulting from the blockage cannot be explained away by the likely change in the composition of readers.

The implications of our findings are not trivial. For democracies, prior studies have shown that news content affects the beliefs and attitudes of citizens and can influence their decisions.[5] In particular, mass media greatly influence the public's views on foreign countries (Huang, Cook, and Xie 2021). If citizens in democratic countries fail to recognize that they are exposed to compromised reporting on foreign regimes, citizens may act and vote in less informed ways. This less apparent mechanism has drawn less attention in public discourse and academic studies than have the disinformation campaigns directly waged by foreign governments.

For authoritarian economic powers, our findings underscore the dilemma of accommodating foreign media. On the one hand, it is legitimate, from the point of view of the regime, to worry about foreign media's influence on citizens' information diet (Chen and Yang 2019).[6] On the other hand, authoritarian regimes, eager to bolster their image overseas, lose the strings that they can pull when foreign media are shut out of their market completely.

Literature Review. Our study contributes to the vast literature in economics examining the determinants of news coverage (cf. an excellent survey by Prat and Strömberg (2013)). Recent examples include analyses by Groseclose and Milyo (2005), Gentzkow and Shapiro ((2006) and (2010)) and Larcinese, Puglisi, and Snyder (2011).

In particular, our paper adds to studies of the influence of governments on news media. Existing research has focused on the role of domestic governments. For instance, Besley and Prat (2006) show that governments may use direct or indirect financial incentives to suppress news.[7] McMillan and Zoido (2004) provide evidence from Peru consistent with the direct channel. Di Tella and Franceschelli (2011) study the media market in Argentina and document that the government uses indirect channels such as government advertising to reduce negative coverage of government misconduct. Gentzkow, Petek, Shapiro, and Sinkinson (2015) show that party control of state governments did not influence the operations of partisan daily newspapers in the period from 1869 to 1928, while Qian and Yanagizawa-Drott (2017), analyzing the patterns of news coverage of human rights abuses in foreign countries, find that the Reagan and Bush Sr. administrations indeed influenced media outlets.

[5] DellaVigna and Kaplan (2007) document the impacts of exposure to news reporting by Fox News. Enikolopov, Petrova, and Zhuravskaya (2011) show that access to independent news sources changed voting behaviors in Russia. La Ferrara, Chong, and Duryea (2012) even show the impact of exposure to soap operas on fertility choices. Several other prominent studies on this issue include those of Strömberg (2004), Gentzkow and Shapiro (2004), Gentzkow (2006), Gerber, Karlan, and Bergan (2009), and Prat (2018).

[6] Chen and Yang (2019) design an experiment in which Chinese students were incentivized to consume news from the New York Times and study such consumption's influence on beliefs of the experiment's participants.

[7] Economic leverage is also wielded by private enterprises to pressure news media to curtail unfavorable reporting about them. Germano and Meier (2013) theorize about this self-censorship mechanism of news media. On the empirical side, Beattie, Durante, Knight, and Sen (2021) show that auto manufacturer recalls are less extensively covered by newspapers in which the firms advertise more regularly.

Other mechanisms have been studied in non-US contexts. Stanig (2015) documents the impact of the defamation law wielded by Mexican governments on news media. Durante and Knight (2012) provide evidence that the news content offered by the public television corporation in Italy shifted to the right when the elected government was center-right. Simonov and Rao (2022) show that an authoritarian government can influence the ideological beliefs of citizens by investing in quality of the government-controlled media platform and non-political news content. Our paper shows that authoritarian governments could also influence news businesses based in democratic countries.[8]

Furthermore, our study contributes to a growing literature in economics and political science that takes advantage of state-of-the-art techniques in computational linguistics. Among prominent examples of related studies, Gentzkow and Shapiro (2010) construct a media slant index based on partisan language used by the media. Shapiro, Sudhof, and Wilson (2020) provide a new sentiment-scoring model that accurately measures sentiment in economic news. Our paper applies the word embedding approach to construct a measure of the negativity of news articles that cover a broad range of news events. Specifically, we utilize an algorithm proposed by Rheault, Beelen, Cochrane, and Hirst (2016) that measures the tone of parliamentary speeches in the UK. Gentzkow, Kelly, and Taddy (2019) stress that such "approaches, us[ing] embeddings as the basis for mathematical analysis of text, can play a role in the next generation of text-as-data applications in social science." Our current study is one example of this research direction.

Hansen, McMahon, and Prat (2018) and Catalinac (2016) both apply topic modeling, LDA in particular, to study political economy issues. Part of our analysis also relies on topic modeling to uncover the underlying themes in the news corpus so that our definitions of various news topics are not excessively arbitrary.

2. Background and Research Questions

2.1. Media Environment in China and the 2019 Crackdown

An increasingly popular narrative about the changing political environment of China runs as follows: The high-water mark of China's opening and liberalization was its entry into the World Trade Organization in 2002. Subsequent decades have seen a plateauing of reforms intended to increase personal freedoms, and many such initia-

[8] Our study is also related to research on how access to news sources can distort news coverage. Ozerturk (2020) theorizes about how access to politicians or governments may be used by these sources to extract more favorable press coverage, and Dyck and Zingales (2003) provide evidence. The mechanism studied in our paper differs in that news outlets compromise their reporting to maintain access to a market for their products.

tives started to reverse course in the 2010s.[9]

As part of this recent trend, the environment in which news media operate in China has deteriorated drastically. The government has started to take more aggressive and preemptive measures to police the internet. It is estimated that in 2020, the total spending on internet censorship in China exceeded 6.6 billion USD.[10] Censors have not only routinely deleted sensitive content online but also blocked entire websites of news media outlets on punitive or even preemptive grounds. One example is the New York Times, which was blocked in 2012 after reporting on the enormous fortunes amassed by relatives of top CCP leaders and has remained inaccessible within China ever since. The Foreign Correspondents' Club of China (FCCC) released a statement on October 22, 2019 regarding the deteriorating environment faced by foreign media in China: "The Great Firewall bars internet users in China from viewing the publicly available websites of 23% of 215 international news organizations with journalists based in China. Among news organizations that publish primarily in English, the most widely spoken foreign language in China, 31% are blocked." Given the restrictions and limitations, a minority of Chinese readers can still access blocked websites with VPNs to bypass censorship. However, according to Freedom House's 2019 report (China), "the government has intensified its restrictions on these tools since new regulations in 2017 placed a ban on the use of unlicensed VPNs."[11]

One dramatic episode is the "rectification" campaign that China launched to clean up its internet in May 2019 (which likely triggered the aforementioned FCCC investigation). Reuters released a detailed news report on this event in early June 2019 and highlighted the scale of this campaign.[12] Numerous news websites and social network accounts were blocked or closed. Many of those casualties, such as Wallstreetcn.com (an influential Chinese financial news publication unrelated to the Wall Street Journal), were publishing materials not even remotely relevant to politics or, as in the case of Wikipedia, were not even news providers. A batch of Western news outlets with considerable coverage of and readership in China were blocked, including not only US- and UK-based news organizations, such as the Washington Post and the Guardian, but also major newspapers and TV programs from Germany, Australia and Singapore.[13]

[9]One example of this view was delivered by Matthew Pottinger in a policy speech on October 23, 2020, "The Importance of Being Candid: On China's Relationship with the Rest of the World. "

[10]See "Buying Silence: The Price of Internet Censorship in China", Jamestown Foundation.

[11]In the same report, the tightening control over using VPNs is also discussed: "VPN providers have noticed growing technical sophistication in the VPN blocking incidents of the past year. Hundreds of VPN services have been banned since 2017 ..." See "Freedom on the net 2019" for more information.

[12]"China launches new internet cleanup campaign; more websites blocked", Reuters, June 12, 2019.

[13]"China blocks websites of major German news outlets", World Association of News Publishers, July 12, 2019.

2.2. A Moving Red Line: US-China Trade Talks Upended

The Chinese government was elusive about the motivations behind this sweeping campaign.[14] Foreign journalists suspected that as the timing coincided with the 30th anniversary of the Tiananmen Incident, this campaign was a preemptive measure to prevent Chinese readers from accessing the inevitable coverage of this event.[15] However, this reason is not sufficient to explain the scale of the campaign and the shutdown of some domestically operated media that would not report on any related sensitive materials.

A more likely reason for the crackdown was to control information on the unexpected breakdown of trade negotiations between the US and China. The prolonged trade talks showed promising signs at the end of April 2019, when a draft trade agreement was crafted in high-level trade talks but took an abrupt turn on May 3, when the US negotiation team reported to "Washington [that] Beijing [had backtracked] on almost all aspects of the draft trade pact."[16] President Trump responded by escalating the trade war, increasing tariffs on US$200 billion worth of Chinese products from 10% to 25%, effective from May 10.

Although the upended trade deal itself was eventually made known to Chinese citizens through official Chinese media, the causes and potential consequences became sensitive. Speculations about the disagreement among top Chinese leaders, the likely miscalculation of Trump's willingness to sign a deal, and the rising economic uncertainty resulting from the worsening Sino-US relationship were all potentially damaging to social stability desired by the Chinese state.[17] The topic of the trade war quietly became a new red line for media without even being noticed by the community of foreign journalists in China. Somewhat later, the true intention of the crackdown was revealed and discussed in Hong Kong-based media.[18]

A quick look at the online search intensity for various topics shows the relevance

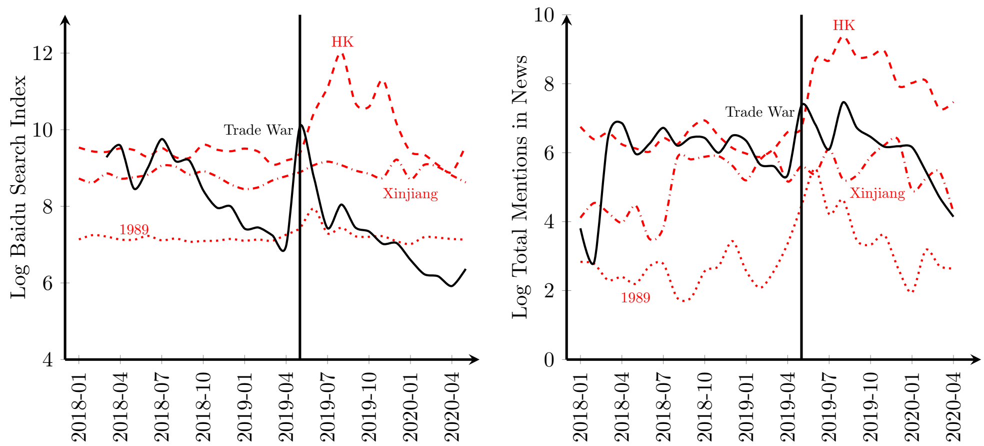
[14] The state-run news agency Xinhua claimed that it was to punish and expose websites for their "illegal and criminal actions" and for failing to "fulfill their obligation to take safety measures or the theft of personal information", according to the Reuters report mentioned earlier.

[15] "China adds Washington Post, Guardian to 'Great Firewall' blacklist (June 9, 2019, The Washington Post)"; "Chinese government blocks Guardian website (June 7, 2019, The Guardian)".

[16] For a summary of the key events of the trade negotiations, see "Timeline: Key dates in the US-China trade war" (January 15, 2020, Reuters).

[17] For media discussion of the causes and consequences of the breakdown of the trade talks, see "How Xi's Last-Minute Switch on U.S.-China Trade Deal Upended It" (May 16, 2019, The New York Times) and "As China Trade Talks Stall, Xi Faces a Dilemma: Fold? Or Double Down?"(May 9, 2019, The New York Times).

[18] For example, South China Morning Post reported on July 9, 2019 that "China's government mulls special stake in wallstreetcn.com as it looks to control the flow of information on trade, economics". It revealed that one alleged crime of the Chinese media outlet wallstreetcn.com, which led to its shutdown, was that it translated the Trump's tweet threatening an increase in tariffs on May 5, 2019, following the upended trade talks.



(a) Baidu Search Index (b) Mentions in News

Figure 1. *Baidu Search Index and Total Mentions in News by News Issue and by Month. Panel (a) illustrates that the number of "trade war" searches on Baidu surged in early May 2019, when the trade deal between the US and China was upended. People indeed searched for "1989" more often in early June 2019. Searches for "Hong Kong" increased dramatically when the situation in Hong Kong intensified in early August. Searches for "Xinjiang" were relatively stable over this data period. Panel (b) illustrates the total mentions by month of those news issues in our news sample over the data period. The search behavior and media coverage intensity are fairly consistent, indicating that both can be driven by events.*

of the fallout of trade negotiations to the media crackdown in terms of timing. We use the Baidu search index—the Chinese counterpart of Google Trends—to proxy Chinese people's attention and display in Figure 1(a) that shows the trends of this index for topics that could represent possible triggers of the crackdown, including the trade war, 1989 (the Tiananmen Incident), Hong Kong, and Xinjiang.[19] The spike in attention to the trade war in May 2019 coincides with the media crackdown, while attention to other news issues peaked at other times or remained flat. The intensity of media coverage of these issues is also consistent with Chinese people's searching behavior and patterns of attention.

We also count the total number of mentions of these keywords (i.e., trade war, 1989, Hong Kong, and Xinjiang) in our news sample (we elaborate on its construction in section 3.1) and display their trends in Figure 1(b). It is highly likely that both media coverage and Chinese people's attention were simultaneously driven by the same set of events. The need to suppress the spiking supply of and demand for news reports on the trade war is consistent with the unprecedented scale of the media crackdown.

[19]On the Baidu search engine, the keyword "Tiananmen" is less informative than "1989" for the Tiananmen Incident, given that the location itself is also a site for military parades and tourism. The Baidu search index results for the keyword "Tiananmen" remained stable in the period until early October 2019, when they surged dramatically. This timing coincides with the military parade for the 70th anniversary of the People's Republic of China.

2.3. Do Foreign Media Value Their Presence in China?

While market access offers effective leverage, it is not the only weapon the Chinese government has to influence foreign media reporting. News organizations, increasingly owned by conglomerates (DellaVigna and Hermle 2017), may have other commercial interests in China. In addition, obstructing foreign journalists and preventing them from accessing news sources is a common tool.[20] Our study focuses on the role of market access for two reasons. First, market access can be measured accurately, while it is difficult to systematically document the business ties and journalist experiences of each media outlet. Second, market access plays an important role in a media outlet's calculations. Despite all of these alleged restrictions and difficulties imposed by the authorities, many mainstream media have made enormous efforts to develop business and cultivate readership in China. For example, the New York Times, the Wall Street Journal, the Washington Post, and Reuters as well as the Guardian have gone out of their way to establish Chinese versions of their websites or translate their news to make them easily accessible to Chinese readers.

The media value access to the Chinese market not only because their presence in China itself brings commercial benefits but also because it could plant seeds of future influence and financial rewards when the political climate changes—a common view shared in the circle of news producers. For example, Craig Smith, a former New York Times's Shanghai bureau chief and China managing director, once stated this calculation explicitly, reflecting on the situation prior to the outlet's 2012 blockage:

> "Our traffic ... grew nearly 70 percent last year alone. The New York Times brand now has a firm foothold in the country and among the global Chinese diaspora. When news media restrictions relax, and I believe they eventually will, the Times's Chinese audience will most certainly take off."[21]

In foreign media outlets' pursuit of profit and influence in an environment where the authority has the means to retaliate for unfriendly reporting, do they intentionally deviate from journalistic standards to avoid repercussions? The anecdote about Bloomberg's editor-in-chief cited in the introduction suggested that the answer is yes but did not explain how. In fact, news media have several degrees of freedom in presenting their news products. They can either interfere with the content of opinion articles or heavily edit hard news. Or they can choose how to manage the reporting of sensitive issues: only toning down negative information or suppressing it entirely. In this paper, we intend to examine these aspects of news reporting on China.

[20]See "Access Denied: Surveillance, harassment and intimidation as reporting conditions in China deteriorate" (December 2017, FCCC).

[21]See "The New York Times vs. the 'Great Firewall' of China (March 31, 2017, The New York Times)."

3. Data

3.1. Sample Construction

We focus on the period from January 2018 to May 2020 to allow a sufficiently long period before and after the media crackdown in June 2019. Our sample is constructed using relevant articles from 19 major news outlets in the US and the UK. The news websites (publishing in English) blocked during the 2019 crackdown include those of the Washington Post, NBC News, the Huffington Post, Breitbart News, the Guardian, and the Daily Mail, which have been inaccessible from mainland China since then.[22] These outlets constitute our treatment group.

As the blocked outlets have either wide circulation or a salient presence in political discourse, our strategy in constructing the control group is to include all the major English-language news outlets with the largest circulations or strong influence, provided that their access status did not change between January 2018 and May 2020. First, we include the top 10 most widely circulated newspapers (except the Washington Post, which is in the treatment group), namely, the New York Times, the Wall Street Journal, Boston Globe, Chicago Tribune, Los Angeles Times, News-Day, New York Post, the Star Tribune, and USA Today.[23] Second, we include influential regional newspapers of similar size in the control group, such as the San Francisco Chronicle, Miami Herald and Dallas Morning News (see Baker, Bloom, and Davis 2016). Third, as the Guardian and Daily Mail are UK-based, we include Reuters (a UK-based international news provider) in the control group to balance the geographical representativeness. In total, there are 13 news outlets in the control group. Table 1 lists the outlets in both groups. Among those in the control group, the New York Times, Reuters and the Wall Street Journal were blocked long before 2018, and their blockage status did not change during the period we examine. For convenience, we label them "always-blocked outlets". The rest of the control group remained unblocked until the end of our data period. We label these "never-blocked outlets".

As discussed in section 2.1, the large-scale crackdown of 2019 was likely to be influence-based. To corroborate this idea, we utilize the Baidu search index of each outlet's name to proxy its influence or potential readership in China. We collect nationwide search intensity data for the name of each media outlet in our sample by

[22] We exclude news sites blocked during this campaign that are based outside the US and the UK such as the Straits Times of Singapore. We verified the blocked status using information released by GreatFire.org, a nongovernmental organization that the FCCC partnered with to analyze and investigate foreign media access in China (discussed in section 2.1). Several independent testing services, such as Chinese Firewall Test, can verify the access status from China for any website.

[23] See the ranking of Cision Media Research, January 04, 2019.

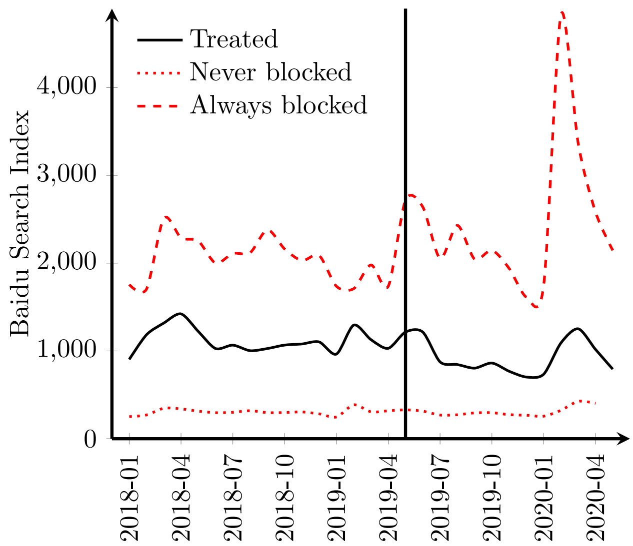


Figure 2. *Average Baidu Search Index by group. Chinese internet users search for the names of always-blocked media outlets most often, even though the outlets have been blocked. The media outlets newly blocked during the 2019 crackdown were searched for more often than the never-blocked outlets. The index for each group increased in February 2020, likely indicating that people searched for foreign media-reported information about the COVID-19 pandemic.*

month and compute the average for each group. Figure 2 illustrates the average index by group. Although always-blocked media outlets remained inaccessible, their names were most often searched for by Chinese internet users. The media outlets blocked during the 2019 crackdown were searched for more often than the never-blocked ones.

Focusing on news articles about China, we scraped from the sample outlets all articles that contained our China-related keywords (i.e., China, Chinese, Hong Kong, Hong Kongese and Hong Konger(s)) at least once. To eliminate articles with irrelevant content, we define and construct a China sample based on the following criteria: 1) we include articles with China-related keywords in the headlines; 2) we include articles that have no country (or people's) names in the headline yet mention China-related keywords at least 5 times in the text; 3) we exclude articles with only the names of other countries and people in the headlines (but no China-related keywords), and 4) we exclude articles categorized into sections explicitly labeled with names of other parts of the world (e.g., "Middle East", "Europe" or "India"). Our main analysis is based on this sample. It is likely that we either exclude articles about China or include articles not about China by setting the threshold of keyword mentions to 5. Therefore, we construct two other samples for robustness checks: the "large sample," in which the threshold for criterion (2) is set to 3 so that we are less likely to exclude articles about China, and the "small sample," in which it is set to 10 so that we are less likely to include articles not about China.

The sample articles fall into three broad and mutually exclusive categories: news,

Table 1. *News Outlets*

Treatment	Control
Breitbart News	The New York Times* #3, blocked by 2012
Daily Mail	Reuters*: blocked by 2015
The Guardian*	The Wall Street Journal* #2, blocked by 2018
Huffington Post	The Boston Globe #10
NBC News	Chicago Tribune #9
The Washington Post* #6	The Dallas Morning News
	Los Angeles Times #5
	Miami Herald
	Newsday #8
	New York Post #4
	San Francisco Chronicle
	Star Tribune #7
	USA Today #1

Note: Newspapers with * have Chinese websites (or regular translations).

opinions, and miscellaneous. The categories can be identified from the sections into which each news outlet classifies the articles. The news category consists of news reports with either objective and descriptive content or investigative and analytical content. The opinion category contains articles including opinions, commentaries, etc. that express opinions of columnists, opinion writers, readers or others. The miscellaneous category includes articles related to arts, entertainment, sports, lifestyle, and a variety of other subjects. Given the diversity of topics in this category, we leave it out of our analysis.

Our main analysis is based on the Chinese sample in the news category, which we label the "news sample" hereafter. This sample is divided into six panels, namely Asia, business, energy (and environment), general (uncategorized) news, politics, and world, based on section titles.[24] The top part of Table 2 shows the number of articles in each panel of the news category in the treatment and control groups separately. Overall, the treatment group contains 13,269 articles, nearly 60% of which are in the control group. The counts' ratio between the treatment and control groups varies across panels partly because the criteria for classification differ by outlet. For example, some outlets may not have a general news section, while others may not have a world section. The same news report on China's environmental protection could fall into the Asia, general news, or politics panel, depending on the outlet. The lack of consistent classification across outlets prevents us from comparing similar panels across outlets

[24]Editorials produced by news staff most likely reflect opinions rather than facts. We leave these out of our analysis. However, the number of editorials about China during the period under investigation was rather small (fewer than 30 in total) relative to the constructed news sample, and inclusion of them does not change any of the results.

Table 2. *Category and panel*

	Treatment	Control	Total
Asia	1512 (4.24%)	1707 (4.78%)	3219 (9.02%)
Business	1582 (4.43%)	9966 (27.92%)	11548 (32.36%)
Energy (and Environment)	95 (0.27%)	1403 (3.93%)	1498 (4.20%)
General News	5764 (16.15%)	1751 (4.91%)	7515 (21.06%)
Politics	2559 (7.17%)	1016 (2.85%)	3575 (10.02%)
World	1757 (4.92%)	6577 (18.43%)	8334 (23.35%)
News (subtotal)	13269 (37.18%)	22420 (62.82%)	35689 (100.00%)
Opinions	1713 (52.40%)	1556 (47.60%)	3269 (100.00%)
Total	14982	23976	38958

but does not affect our conclusion drawn by considering the news sample as a whole.

As a comparison, we also examine China-related articles in the opinion category, which we label "the opinion sample". As shown in Table 2, the number of articles per outlet is 1,713 and 1,556 in the treatment and control groups, respectively, much lower than the counts for the news sample.

3.2. Measuring Negativity towards China

To measure the tone of news articles, we first create a corpus-based sentiment dictionary that assigns emotion or tone scores to each word, and then computes the average score for each article. The procedure is as follows: 1) representing each word in the corpus with a numerical vector (embedding), 2) measuring the emotion or tone of each word using a sentiment lexicon, and 3) aggregating to the article level. This approach is appealing not only because it is unsupervised and requires little human input but also because the vectorization process is domain-specific or adaptive to context: vectors encode the meanings of words and reflect how words are used in the corpus.[25] This feature is particularly relevant for this study: the same word may carry different emotional valences in different contexts (such as parliamentary speeches, Wikipedia, and news media content) or in different time periods in news content.

While Appendix A details the algorithm, training process and construction of emotion at the word and article levels, we outline the key ideas below. First, we create a vector space model, which turns the vocabulary of our news article corpus into numerical vectors following the global vectors for word representation (*GloVe*) algorithm (Pennington, Socher, and Manning 2014). This algorithm explicitly utilizes ratios of word-word co-occurrence probabilities to encode some form of meaning of each word.

[25]Several problems associated with the dictionary-based approach can thereby be avoided; e.g., dictionaries might find it difficult to deal with polysemes and often fail to capture all synonyms.

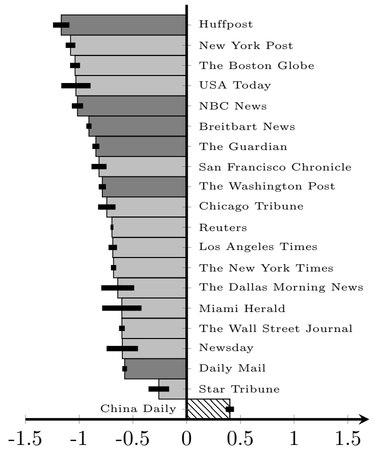
Second, we measure the tone of each word by following the algorithm developed by Rheault, Beelen, Cochrane, and Hirst (2016). The essence of this approach is to compute a given word's similarities with a group of positive seed words and a group of negative seed words and then use the net aggregate distance to represent the focal word's tone. Specifically, the tone s_i of word w_i is calculated as

$$s_i = \sum_{p \in P} \frac{w_i w_p}{||w_i||||w_p||} - \sum_{q \in Q} \frac{w_i w_q}{||w_i||||w_q||},$$

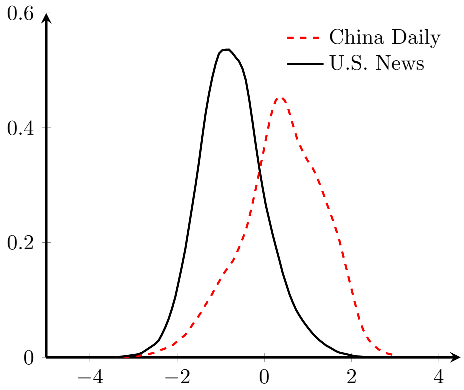
where P is the positive seed word set, Q is the negative seed word set, and $||w_i||$ is the norm of word vector w_i. Note that the dot product of vectors w_i and w_j is the cosine similarity, representing the distance between word vectors i and j. Seed words are chosen so that they have "no multiple, opposite meanings, when used as a specific part of speech, and ... exclude terms with domain-specific meanings" (Rheault, Beelen, Cochrane and Hirst 2016). A larger score s_i implies that word i is more positive in tone.

Third, we aggregate words' tones to the article level and construct four types of measures. Our main measure for the tone of an article is the simple average of all words' scores s_i in each article (after excluding stop words, etc.). One may worry that the simple average score of an entire article may contain excessive noise because the article may comment on China positively but describe the context negatively, or vice versa. To alleviate this concern, we construct the second measure—the China-based score, which is the average score of words that appear only in sentences mentioning China or Chinese. Another concern is that the score of each word may not precisely measure the tone, especially for relatively neutral words with low similarity scores. For robustness tests, we construct a third type of measure—the nonneutral score, which is the average of scores of only words with strong positive or negative emotions, i.e., excluding those with scores within one (or two) standard deviation(s) of the mean score of the entire lexicon. Lastly, Pennington, Socher, and Manning (2014) provide pretrained word vectors resulting from training on a corpus that consists of a large number of Wikipedia articles. To corroborate our training process, we compute the average tone of each news article using those pretrained word vectors. We expect this Wikipedia-based measure to correlate with the other three measures constructed using our news corpus.

Our tone measures can be validated at both the outlet and article levels. We first contrast US and UK media outlets in our sample with China Daily, the Chinese government's mouthpiece. Our premise is that China Daily adopts a more positive tone in China coverage than do our sample outlets. The left panel of Figure 3 illustrates that the average article-level tone score is positive for China Daily and negative for each of our sample news outlets. The right panel of Figure 3 further shows that most of the



(a) Average tone by media outlet



(b) Distribution of article tone

Figure 3. *Validity at the Press Level. The left panel displays the means and confidence intervals of article-level tone scores (the main measure) of UK- and US-based outlets in our news sample and those of China Daily. The right panel shows the article-level score distribution for outlets in our news sample and that for China Daily.*

article-level score distribution for our sample outlets lies to the left of that for China Daily. The pattern revealed in both figures confirms our premise and further supports validity of our measures. We also demonstrate in Appendix A that our tone scores are correlated with the ratings of trained human assistants and provide several samples from New York Times publications for illustration.

3.3. Summary Statistics

Table 3 presents the summary statistics of the variables used in our analysis. Columns (1) and (2) show the means and standard deviations of those variables in the news sample for the treatment and control groups, respectively, while column (3) reports the differences between the means and the standard errors of the differences clustered at the press level. The means of the simple average tone scores are -0.72 and -0.75 for the control and treatment groups, respectively. The China-based tone scores are approximately 0.15 units lower for each group than the simple average scores. Removing relatively neutral words within 1 standard deviation around the means (i.e., using the nonneutral score) lowers the tone scores further by approximately 0.1 units. The Wikipedia-based sentiment scores are even more negative.

It is worth noting that the tone of the treatment media is not significantly more negative than that of the control group. This is unsurprising because the overall media tone is determined by various factors beyond the concern of market access to China.

The articles in the treatment group are longer and mention China more frequently.

Table 3. Summary of Statistics

	News			Opinions		
	Treatment mean (sd)	Control mean (sd)	Diff mean (se)	Treatment mean (sd)	Control mean (sd)	Diff mean (se)
	(1)	(2)	(3)	(4)	(5)	(6)
Default score	-0.75	-0.72	-0.03	-0.74	-0.61	-0.12
	(0.79)	(0.74)	(0.09)	(0.56)	(0.69)	(0.05)
China-based score	-0.91	-0.87	-0.03	-0.90	-0.73	-0.16
	(0.91)	(0.86)	(0.09)	(0.71)	(0.80)	(0.05)
Score excluding 1 std	-1.00	-0.96	-0.04	-0.98	-0.78	-0.19
	(1.22)	(1.19)	(0.14)	(0.83)	(1.02)	(0.07)
Wiki-based score	-1.13	-1.04	-0.09	-1.03	-0.89	-0.14
	(1.10)	(1.03)	(0.09)	(0.78)	(0.96)	(0.07)
Wordcount	1059.77	646.50	413.27	1714.23	902.19	812.04
	(1580.04)	(568.66)	(201)	(1922.28)	(1133.41)	(173)
Share: Wordcount $<$ 25%	0.12	0.33	-0.20	0.19	0.32	-0.13
	(0.33)	(0.47)	(0.09)	(0.39)	(0.47)	(0.07)
Share: 25% $\geq$ Wordcount $<$ 25%	0.25	0.25	0.00	0.22	0.29	-0.07
	(0.43)	(0.43)	(0.04)	(0.41)	(0.45)	(0.03)
Share: 50% $\geq$ Wordcount $<$ 75%	0.30	0.22	0.08	0.24	0.26	-0.01
	(0.46)	(0.41)	(0.03)	(0.43)	(0.44)	(0.05)
Share: Wordcount $\geq$ 75%	0.32	0.21	0.12	0.35	0.14	0.21
	(0.47)	(0.41)	(0.11)	(0.48)	(0.35)	(0.08)
Freq. China & Chinese	14.02	11.26	2.76	13.98	13.50	0.48
	(11.67)	(9.36)	(2.09)	(9.31)	(9.76)	(0.83)
Mention Tian'anmen	0.03	0.02	0.01	0.05	0.08	-0.03
	(0.17)	(0.14)	(0.009)	(0.22)	(0.27)	(0.013)
Mention HK	0.28	0.26	0.02	0.24	0.24	0.00
	(0.45)	(0.44)	(0.05)	(0.43)	(0.43)	(0.02)
Mention COVID	0.24	0.16	0.08	0.16	0.17	-0.01
	(0.43)	(0.37)	(0.02)	(0.36)	(0.37)	(0.02)
Mention trade-war	0.15	0.24	-0.09	0.27	0.21	0.06
	(0.36)	(0.42)	(0.04)	(0.45)	(0.41)	(0.02)
Mention trade	0.34	0.49	-0.15	0.56	0.48	0.08
	(0.47)	(0.50)	(0.06)	(0.50)	(0.50)	(0.03)
Mention corruption (*100)	1.78	1.16	0.62	3.93	2.54	1.39
	(13.23)	(10.72)	(0.51)	(19.43)	(15.73)	(0.91)
Mention leaders' scandal (*100)	0.38	0.14	0.24	0.29	0.00	0.29
	(6.13)	(3.72)	(0.34)	(5.40)	(0.00)	(0.12)

Notes: The standard error in columns 3 and 6 are clustered at the press level.

On average, an article in the treatment group is approximately 1060 words long and mentions China-related keywords 14 times, while the respective counts for the control group are only two-thirds of those of the treatment group. Dividing the news sample by length (i.e., by word count) into four quartiles, we show that both the treatment and control groups have a nontrivial presence in each quartile — the share of articles in each quartile for each group ranges between 12% and 33%. Despite the imbalance, both groups have sufficient observations for articles of different types.

We further examine the likelihood of these articles mentioning specific keywords that pertain to specific news issues. The definition and construction of such keywords can be found in Appendix D. Among articles in the treatment group, 24% mention COVID-19-related keywords, 3% mention Tiananmen and 28% mention Hong Kong-related keywords. The likelihood of mentioning the trade war is 15%, while that of mentioning the word "trade" is 34%. Only 2% of articles mention corruption-related keywords. All of these issues are similarly present in the control group.

Columns (4) and (5) of Table 3 display the means and standard deviations of these variables for the treatment and control groups, respectively, in the opinion sample, while column (6) reports the differences between the means and the standard errors of the differences. The opinion articles published in the treatment outlets are on average more negative toward China and longer than those in the control outlets.

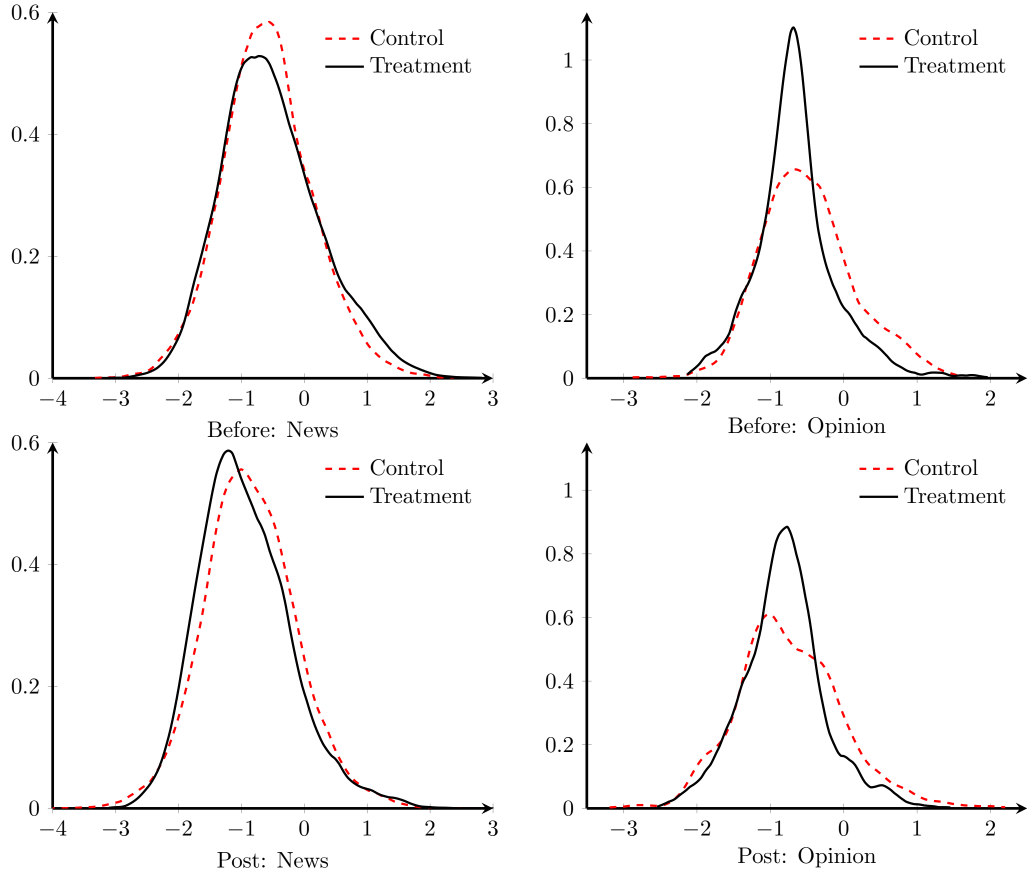
Does the distribution of tones change differently after the blockage across the treatment and control groups? We present the kernel distributions of tone scores for the treatment and control groups before and after the blockage, in Figure 4. Panel (a) on the left illustrates the distributions for the news sample, and panel (b) on the right illustrates the distributions for the opinion sample. As the top-left figure shows, the distribution of the control group is slightly more compressed than that of the treatment group, but the two distributions overlap for the most part and had little visible difference before the crackdown. The bottom-left figure shows a clear leftward deviation of the treatment group's distribution from that of the control group. Most of the treatment group's distribution shifted to the left of that of the control group, suggesting that after the crackdown, the treatment group became more negative in tone than the control group. As to the opinion sample, the distributions of the treatment and control groups have different shapes but overlap to some extent. More importantly, there was no visible shift after the crackdown.

4. Identification Strategy

As discussed in section 2.1, the large-scale crackdown in May 2019 was likely based on the influence of news outlets rather than the content published by specific outlets, and intended to control information on and attention to the unexpected breakdown of trade negotiations. This consideration motivates our use of a difference-in-differences (DID) model to identify how losing access to China affected the media's handling of China-related articles. We start by comparing changes in the tone toward China of the treated outlets with those of the control outlets using the following specification:

$$y_{ipjt} = \beta \left(T_j \times Post \right) + X_i \gamma + \rho_p + \mu_j + \lambda_t + \epsilon_{ipjt} \tag{1}$$

where y_{ipjt} is the measure of the tone of article i in panel p published by outlet j at time (in month) t. T_j is the indicator for the treatment group; $Post$ is a dummy variable that takes the value of 1 if article i was published in or after June 2019 and is 0 otherwise, and X_i is a vector of article-level control variables, including the total word count and the total number of occurrences of words "China" and "Chinese" in article i, which capture article i's length and relevance to China. We include panel, outlet, and month fixed effects— denoted by ρ_p, μ_j and λ_t, respectively—to control for panel-



(a) Kernel Densities of the News Sample (b) Kernel Densities of the Opinion Sample

Figure 4. *Kernel Density. The solid lines represent the distributions of the treatment group, and the dashed lines represent those of the control group. Panel (a) illustrates the contrast between the periods before and after the blockage for the news sample. Panel (b) presents the counterpart for the opinion sample.*

, outlet- and time-specific factors that affect the tone of news articles. The inclusion of these fixed effects renders the dummy variables T_j and $Post$ redundant in this regression. As discussed in section 3.1, the classifications of panels may differ greatly by outlet. Therefore, we also include higher-dimension panel-by-outlet fixed effects in some specifications to control for unobservable characteristics at the panel-by-outlet level. All standard errors are clustered at the press level.

In addition to measures at the intensive margin (i.e., tones), we also explore measures at the extensive margin, i.e., the number of articles over a fixed period of time, in a specification similar to Equation (1) with the controls adjusted accordingly.

The key coefficient of interest is β in Equation (1), which captures the impact of the 2019 blockage on outcome variables. We attribute a significant estimate of β to losing market access under the parallel trends' assumption that the treated media outlets would have followed a trend of the outcome variables parallel to that of the control outlets had they not been blocked in 2019.

The first challenge to our research design is that the number of outlets in our sample is relatively small, especially that of treatment outlets. The within-outlet correlations may lead to an underestimation of standard errors. To address this concern, we report three sets of p values adjusted for this bias. First, we follow the suggestion by Bertrand, Duflo, and Mullainathan (2004) to report the cluster-correlated Huber-White standard errors for all specifications. Second, we report p values computed using the cluster-adjusted wild bootstrap (WB) method, following (MacKinnon and Webb 2018) and considering each press as a cluster. Third, we also report p values based on the randomization inference (RI) test (Rosenbaum 2002).[26] Both WB and RI approaches yield conservative estimates. If the respective p values are sufficiently small, the over-rejection problem caused by the small number of clusters should not be a serious concern.

Another challenge to our design is that the blockage was endogenous to the news content or the preexisting content trends. To address this concern, we first drop all articles that ever mention the suspected triggers of the crackdown, namely the trade war or the Tiananmen Incident, to test the robustness of the result. Next, we test whether the treatment outlets had developed an increasingly harsher tone over time before the blockage compared to the control group using an event study model specified as follows:

$$y_{ipjt} = \Sigma_{\tau=-12}^{11} \alpha_\tau (T_j \times Month_\tau) + X_i \gamma + \rho_p + \mu_j + \lambda_t + \mu_{ipjt}. \tag{2}$$

We treat months from January 2018 to April 2018 as the base period and compare the difference in the outcome variable between the treatment and control groups in the subsequent months with that in the base period. $Month_\tau$ (where $\tau = -12, \ldots, 11$) are dummy variables for the months from May 2018 to April 2020. The value $\tau = 0$ indicates the month of May 2019, when the crackdown occurred. If there is no difference in preexisting trends between the treatment and control groups, we would expect α_τ—the coefficients of the interaction terms between the treatment dummy variable and the month dummy variables $T_j \times Month_\tau$—not to be significantly different from 0 for $\tau < 0$. Additionally, if the blockage made the treated outlets harshen their tones toward China, we would expect that α_τ becomes negative for $\tau > 0$.

One may worry that the estimated blockage effect is confounded by a so-called chilling effect, i.e., that the unblocked media were "scared" into toning down negativity toward China after the crackdown. To address this concern, we reestimate both the DID and event study models using only the always-blocked media outlets as the control group. We further compare the change in tone of the never-blocked media

[26] We construct the sampling distribution of the estimated $\hat{\beta}$ by repeatedly randomly assigning the treatment outlet and estimating the placebo effects. The p value is computed by noting where our estimated effect lies in the distribution of placebo effects.

outlets after the blockage with that of the always-blocked outlets using a DID model otherwise identical to Equation (1) except that the never-blocked media outlets are relabeled as the treatment group, and the always-blocked outlets as the control group. In the presence of the chilling effect, we would expect the estimated effect in this placebo test to be positive. The absence of such an effect would provide us with more confidence in the validity of the construction of the control group.

Another potential threat to our identification strategy is that media outlets specialize in different areas and would have responded differently to newsworthy events occurring after the blockage, particularly those related to authoritarian politics. In this case, the estimated effect would be attributable to the difference in media specialization instead of market access. To mitigate this concern, we first verify the robustness of our results by excluding news articles mentioning prominent news issues occurring after the blockage, such as the Hong Kong protests and the COVID-19 crisis. Second, we restrict our analysis to the treated media outlets and study whether they handled China-related news differently from Russia- and Iran-related news in response to the crackdown. Russia and Iran are chosen as comparison because they also are authoritarian regimes, and the news media pay a considerable amount of attention to political and foreign affairs in the two countries. We scraped Russia- and Iran-related news articles from our sample news outlets and constructed Russia and Iran samples following the same criteria as those for the China sample. Summary statistics of the Russia and Iran samples can be found in Table 16 of Appendix E. We estimate the following DID model with the Russia- and Iran-related news sample as the control group and the China-related news sample as the treatment group:

$$y_{ipct} = \beta_c \left(China_c \times Post \right) + X_i \gamma + \rho_p + \nu_c + \lambda_t + \epsilon_{ipct}, \tag{3}$$

where $China_c$ is an indicator of article i being related to China, and ν_c represents country fixed effects. If β_c in Equation (3) is negative, we can be more confident that the change in tone of the treated media relative to the control media arose from the crackdown in China rather than the treated media's general response to foreign politics.

To further mitigate the confounding bias caused by time-varying group-specific factors, e.g., an overall change in the tone of the treated outlets toward authoritarian regimes, we combine the samples of news on China, Russia and Iran published by the treated and control media and consider a DDD model in which Russia- and Iran-related news articles are used as an additional comparison group. Specifically,

$$\begin{aligned} y_{ipcjt} &= \delta_1 \left(T \times Post \right) + \delta_2 \left(T \times China_c \right) + \delta_3 \left(China_c \times Post \right) \\ &+ \beta_{triple} \left(T \times China_c \times Post \right) + X_i \gamma + \rho_p + \nu_c + \mu_j + \lambda_t + \epsilon_{ipcjt}, \end{aligned} \tag{4}$$

where y_{ipcjt} is the measure of tone for article i in panel p related to country c published by outlet j at time (in month) t. The coefficient β_{triple} captures how the difference in tone toward China between the treated and control media changed after the blockage in comparison to the changes in the difference of tone toward Russia and Iran. A statistically insignificant DDD estimate of β_{triple} would indicate that the DD estimate of the blockage effect arises from the changes in the treated media's dealing with news related to authoritarian regimes in general rather than the impact of losing access to China.

5. Does the Market Access Matter for News Reporting?

5.1. Baseline Results

How did the news outlets change their tone after losing access to the Chinese market? Column (1) of Table 4 shows the results of estimating the baseline DID model without fixed effects but including the main effects. The statistically insignificant group main effect (the coefficient of T) suggests that the treatment and control media did not differ in the tone toward China before the crackdown. The time main effect (the coefficient of *Post*) is negative and statistically significant, suggesting an overall harshening of tone across media outlets. The coefficient of interest is that of the interaction between the treatment and Post dummy variables $T \times Post$, which shows that the average tone score of articles published in treatment outlets decreased by 0.19 (or 0.23 standard deviations) after the blockage relative to that in the control group. The estimated negative effect remains significant at the 1% level after including press and month fixed effects (Column (2)) or the press-by-panel fixed effects (Column (3)).

The blockage impact on the news tone was fairly large. As Table 3 shows, the average sentiment scores of our news sample articles and those in China Daily are approximately -0.75 and 0.44 respectively, with a gap of 1.19. Our estimated blockage effect is approximately 15% of this gap. In other words, the blockage made the treated media outlets' tone deviate from that of China Daily by additional 15% relative to that of the control media outlets.

Next, we consider the opinion sample. Column (4) of Table 4 reports the results with main effects. Interestingly, both group and time main effects are negative and significant, and the estimated coefficient of the interaction $T \times Post$ is statistically insignificant. The result suggests that the treated media outlets tended to be harsher toward China than the control media, and both groups became more negative after the crackdown, but the treatment media did not change differently from the control media outlets after being blocked. The result remains qualitatively the same after me-

Table 4. *Baseline DID result: Tone changes, default tone as outcome variable*

	News Sample			Opinions Sample	
	(1)	(2)	(3)	(4)	(5)
T × Post	-0.194***	-0.180***	-0.172***	0.078	0.059
	(0.066)	(0.052)	(0.049)	(0.086)	(0.059)
[*WB p-value*]	[0.096]	[0.042]	[0.046]	[0.743]	[0.583]
{*RI p-value*}	{0.019}	{0.043}	{0.045}	{0.528}	{0.538}
T	0.144			-0.181***	
	(0.100)			(0.047)	
Post	-0.240***			-0.252***	
	(0.037)			(0.050)	
Controls	Yes	Yes	Yes	Yes	Yes
Month FE	No	Yes	Yes	No	Yes
Press FE	No	Yes	No	No	Yes
Panel FE	Yes	Yes	No	No	No
Panel × Press FE	No	No	Yes	No	No
R-Squared	0.099	0.140	0.153	0.041	0.131
N	35,689	35,689	35,685	3,269	3,269

Notes: Controls include the total word count and the total occurrences of the word "China" and "Chinese" in the article. Standard errors in parentheses clustered at the media outlet level; * p<0.1, ** p<0.05, *** p<0.01. *P*-values computed with wild bootstrap and randomization inference are reported in the square and curly braces respectively.

dia and month fixed effects are included (column (5)).

To examine whether our estimate is subject to the over-rejecting problem caused by the small number of clusters, we show the p values of the effect computed using cluster-adjusted wild bootstrap (WB) and randomization inference (RI) in the square and curly braces, respectively, for each specification in Table 4. For the news sample, the WB-based p values are 9.6%, 4.2% and 4.6% for specifications with main effects, press fixed effects, and press-by-panel fixed effects, respectively. The RI-based p values are even smaller. It is worth noting that WB and RI at the press level likely lead to an under-rejection problem here because of the small number of the treatment media outlets and the heterogeneity of control media. Despite the under-rejection possibility, all p values are still below 10%, which strongly corroborates the robustness of our result. To further eliminate the possibility that a particular outlet drives our findings, we reestimate Equation (1) by excluding one media outlet at a time. The result, reported in Table 15 of Appendix C, remains robust. In contrast, the WB- or RI-based p values of estimates using the opinion sample are higher than 50%, confirming no significant blockage effect in this sample.

The contrast between the news and the opinion samples is striking but intuitive. It has long been a practice and a tenet in journalism that there is a "wall" between the news and opinion sides of business; i.e., reporters working for the news section and those working for opinion sections remain independent. The views expressed in opinion articles typically belong to the writers, not the outlets; therefore, the outlets do not claim responsibility for those views. Our results suggest that the media compromise their news production, for which they claim responsibility, but do not interfere with their opinion publications, for which they do not. We thus focus on the news sample in the rest of this paper.

5.2. Robustness Tests

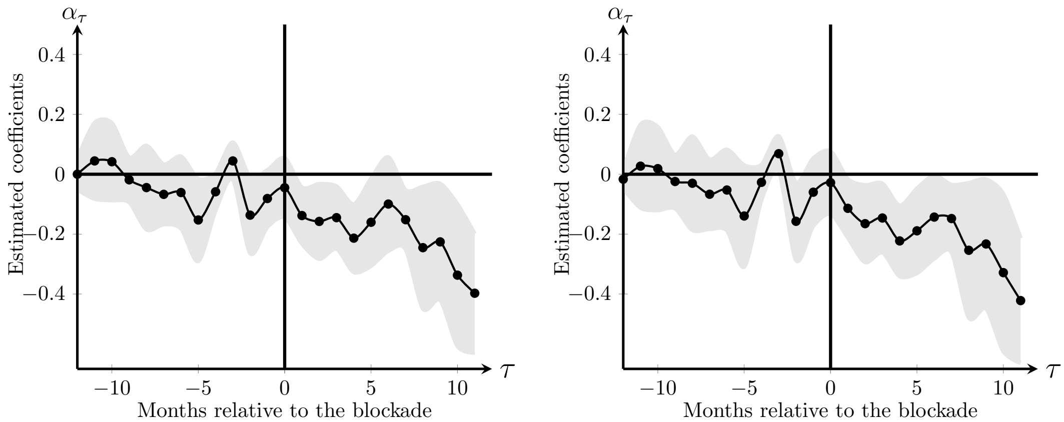
As mentioned in Section 4, several concerns may remain regarding the validity of the identification strategy and the robustness of the result. We will examine them in this subsection.

Crackdown endogenous to news content? To examine whether the crackdown was endogenous to news content, we first investigate whether news articles that mention the trade war and/or Tian'anmen drive the identified results. We reestimate Equation (1) by excluding articles that *ever* mention the following terms one-by-one: "trade war" , "trade" , "Tiananmen" , and either "Tiananmen" or "trade war." The respective results are reported in columns (1)-(4) of Table 10 in Appendix B. Note that removing articles that ever mention "trade" leads to discarding approximately 40% of the sample. Nevertheless, the identified blockage effects on the news tone remain robust, and the magnitude is similar to the baseline estimate in Table 4 for all specifications. The WB- and RI-based p values of the estimated blockage effects are below or slightly above 5%, further reassuring us that our main result is not driven by the suspected triggers of the crackdown.

Preexisting trends in news content? We use the event study model to examine the time at which the trends in tones in the treatment and control groups diverged. We estimate Equation (2) using our benchmark tone scores as the outcome variable. Figure 5(a) illustrates the estimated coefficients α_τ (versus the number of months relative to the blockage) and their 95% confidence intervals.

Except for α_{-5} and α_{-2} that are marginally significant, the estimated coefficients α_τ are overall statistically insignificant for $\tau < 0$, indicating no difference in pre-trends between the treatment and control groups before the blockage. This finding rules out the concern that the treated outlets were blocked in May 2019 because they exhibited an increasingly negative tone toward China.

In contrast, starting from June 2019 (the month immediately after the blockage), the estimated coefficients α_τ are consistently negative and significant with only one exception, namely α_6. In other words, articles published by the treated media outlets exhibited a greater deterioration in tone than that observed for articles in the control group. The timing of this divergence coincides precisely with the crackdown waged by the



(a) Control Group I: All (b) Control Group II: Always-blocked

Figure 5. *Event Study Model. The left panel (a) illustrates coefficients and the associated confidence intervals estimated with the event study model and by using all outlets in the control group. The right panel (b) illustrates the respective coefficients resulting from using the always-blocked outlets as the control group, i.e., control group II. The patterns in both estimations are rather similar. There is no difference in the preexisting trends between the treatment and control groups before the blockage. The timing of the divergence between the treatment and control groups coincides precisely with the crackdown. The period between January 2018 and April 2018 is treated as the base period. Month*$_\tau$ *(where* $\tau = -12, \ldots, 11$*) represents dummy variables for the months from May 2018 to April 2020. In particular,* $\tau = 0$ *indicates the month of May 2019, when the crackdown occurred.*

Chinese government, suggesting that the effect arises from the response of treated outlets to the blockage.

Chilling Effects? Does our result arise because the never-blocked outlets in the control group responded to the crackdown by practicing more self-censorship? To explore this, we reestimate the same event study model of Equation (2) using only always-blocked outlets as the control group. The pattern, illustrated in Figure 5(b), is rather similar to that for the entire control group shown in Figure 5(a), indicating that it is not driven by a potential chilling effect.

We further test whether the never-blocked media outlets responded to the crackdown differently from always-blocked outlets, which did not respond. We perform a placebo test by relabeling the always-blocked media as the control group, and the never-blocked media as the pseudo-treatment group. Using the sample for only these two groups of media outlets, we estimate Equation (1) for a variety of measures of news tone and observe no significant blockage impact on the never-blocked media. The result, shown in Table 11 in Appendix B, reassures us that there was no significant chilling effect and that our construction of the control group is valid.

Different responsiveness to post-crackdown events? Could the harsher tone have arisen because the treated outlets by nature were more responsive to prominent newsworthy events occurring after the blockage? Specifically, media outlets may have exhibited inherently different responses to the most salient China-related news stories, namely the 2019 pro-democracy protests in Hong Kong and the COVID-19 pandemic.[27] To address this concern, we estimate Equation (1) by excluding, in separate analyses, articles that *ever* mention any Hong Kong-related keyword, COVID-19-related keywords, and either Hong Kong- or COVID-19-related keywords (see Appendix D for details of keywords). The result remains statistically significant. To save space, we present this in Appendix B. The small WB- and RI-based p values of the estimated blockage effects, shown respectively in the square and curly braces in each column of Table 12, provide reassuring evidence that our result is not driven by the coverage of these two topics.

Different responsiveness to authoritarian politics? Another likely threat to the validity of identification is that the treated media differ from the control media in their potential responsiveness to issues related to authoritarian politics or foreign affairs. To address this concern, we restrict sample outlets to the treated media, and use China-related news articles as the treatment group, and Russia- and Iran-related news articles as the control group to estimate Equation (3). The results with main effects and fixed effects are reported, respectively, in Columns (1) and (2) of Table 5. Interestingly, the estimated main effects in column (1) reveal that the treated media in fact had adopted a more positive tone toward China than toward Russia and Iran before the blockage and became more negative toward the latter over time. More importantly, the coefficients of the interaction between the indicator for China-related articles and the Post dummy variable ($China_c \times Post$) are significantly negative, showing that the treated outlets raised the negativity in tone toward China rather than toward Russia and Iran after the crackdown. Our finding suggests that the change in tone toward China was not driven by the potential difference in the reporting focus between the treated and control media.

One may still worry that the increased hostility toward China is part of a general trend of changes in attitude toward authoritarian countries among the media, which could confound our DID estimate of the blockage effect. To explore this, we estimate the DDD model (4) with the China, Russia and Iran samples combined. The results with main effects and fixed effects are reported, respectively, in columns (3) and (4) of Table 5. The significant and negative coefficients of the triple interactions $T \times China_c \times Post$ show that the difference in negativity of tone toward China between the treated and control media became larger after the blockage than the difference in negativity

[27]Among the thirteen news topics that we identify using the topic model (as discussed in detail in section 6.1), the Hong Kong protests and the COVID-19 pandemic are the only two news topics that became relevant after the crackdown. The Hong Kong protests started gaining momentum in the middle of June 2019 and lasted approximately 7 months, waning after early January 2020. The COVID-19 pandemic started in January 2020 and continued throughout the entire year 2021.

Table 5. Russia and Iran samples as a comparison group

	Treatment Media with China, Russia and Iran Samples		All Media with China, Russia and Iran Samples	
	Difference in Differences		Triple Differences	
	(1)	(2)	(3)	(4)
China ×Post	-0.303***	-0.367***	-0.063**	-0.159***
	(0.064)	(0.051)	(0.023)	(0.032)
China	0.632***		0.608***	
	(0.085)		(0.024)	
Post	-0.147***		-0.182***	
	(0.024)		(0.021)	
T			0.158*	
			(0.085)	
China × T × Post			-0.233***	-0.233***
			(0.070)	(0.076)
[*WB p-value*]			[0.036]	[0.031]
{*RI p-value*}			{0.056}	{0.061}
T × Post			0.031	0.040
			(0.033)	(0.026)
T× China			0.004	0.002
			(0.096)	(0.105)
Controls	Yes	Yes	Yes	Yes
Press FE	No	Yes	No	Yes
Month FE	No	Yes	No	Yes
Panel FE	Yes	Yes	Yes	Yes
Country FE	No	Yes	No	Yes
R-square	0.177	0.223	0.233	0.285
N	18992	18992	59223	59223

Notes: Controls include the total word count and the total occurrences of the word "China" and "Chinese" in the article. Standard errors in parentheses clustered at the media outlet level; * p<0.1, ** p<0.05, *** p<0.01. P-values computed with wild bootstrap and randomization inference are reported in the square and curly braces respectively.

toward Russia and Iran. The WB- and RI-based p values of the estimated coefficient of this triple interaction are close to 5% for both specifications. It is worth noting that the main effect on the China dummy variable shows that the control media are friendlier toward China than toward Russia and Iran (column (3)). The insignificant coefficient of the interaction $T \times China_c$ shows that the treated media were not particularly harsh toward China before the blockage (columns (3) and (4)). The insignificant coefficient of the interaction $T \times Post$ shows that the treated media's tone toward Russia and Iran did not change differently from that of the control media. In summary, while all media outlets indeed became increasingly negative toward the three authoritarian regimes, the treated media became additionally harsh toward China after the blockage.

Robustness to alternative measures and samples. We estimate Equation (1) with alternative measures discussed in section 3.2 to assess the robustness of the result. Table 13 of Appendix B reports separately the results obtained using the China-based scores, the nonneutral scores and the Wikipedia-based scores. These estimates, although varying in magnitude and level of significance, are consistent with our baseline result (column (2) of Table 4). Next, we use tone scores constructed from word representation models with various dimensionality parameters. The results (reported in Table 14 of Appendix B) are consistent with each other and corroborate the main result. Lastly, we estimate the DID model (1) using two alternative samples, namely the large sample and the small sample as discussed in section 3.1). The estimates, also reported in Table 13 of Appendix B, are close to those obtained using the default news sample (column (2) of Table 4), suggesting that our results are robust to the choice of sample.

6. Appeasing Censors: How Not to Anger China?

The crackdown removed market access for the treated media and in doing so erased a constraint on their reporting. Having been kicked out, they may have become less worried about upsetting Chinese censors. It is interesting to identify the news topics for which the media outlets adjusted their reporting strategy, and in particular, to explore whether the adjustment was more salient for topics that might annoy Chinese censors. To this end, we use topic modeling to discover the topics underlying the news reports and then examine the impacts of the blockage on each topic at both intensive and extensive margins.

6.1. Intensive Margin: News Tone across Topics

To endogenously characterize topics or themes, we estimate an LDA topic model (Blei, Ng, and Jordan 2003) with our China news corpus. LDA is a generative probabilistic model in which the assignment of words to topics and the assignment of topics to documents are jointly estimated. In this model, a topic is defined as a distribution over words; i.e., word probabilities for a given topic sum to one. A document is a distribution over topics; i.e., the topic proportions across all topics for a document sum to one. LDA trades off two goals: (i) for each document, the algorithm allocates words to as few topics as possible, and (ii) for each topic, the algorithm assigns a high probability to as few words as possible. Therefore, topics (weighted word lists) emerge endogenously from the estimation without requiring pre-specified words to characterize the topics. Another output is a multinomial distribution over topics for each document (weighted topic lists). The details of our estimation are relegated to Appendix F.

The number of topics K is the key choice to make; it varies based on the study's purpose. For example, choosing a large number of topics, we obtain topics such as China's relations with Japan, Europe and the UK. Choosing a smaller number of top-

(a) Topic 1 (b) Topic 4 (c) Topic 6

Figure 6. *Example Word Clouds. Panels (a), (b) and (c) show word clouds for the news topics of the trade war, economic growth and human rights, respectively.*

ics, we obtain coarser topics such as China's foreign relations.

We experiment with different numbers of topics and set $K = 13$ in the benchmark model. The general rule is that we choose the number of topics so that several key topics relevant to our analysis, such as human rights, the trade war, and growth, become distinct and so that those topics are not repetitive.[28] All thirteen topics identified are clearly interpretable: the trade war (topic 1), energy (topic 2), industry (topic 3), economic growth (topic 4), financial markets (topic 5), human rights (topic 6), Huawei and high-tech security (topic 7), relations with the US (topic 8), relations with the UK/Australia (topic 9), Taiwan, North Korea and the South China Sea (topic 10), social issues (topic 11), Hong Kong anti-extradition bill protests (topic 12), and COVID-19 (topic 13). Tables 17 and 18 in Appendix F present the top 20 keywords for each news topic, which provide a foundation for our interpretation. All the news topics are salient and have received considerable coverage. Most topics (except the COVID-19 crisis and Hong Kong protests, which occurred after the crackdown) are recurring topics covered both before and after the crackdown. We illustrate with word clouds three example topics, namely the trade war, economic growth and human rights in Figure 6, and a full list of word clouds is presented in Figure 12 of Appendix F.

Based on the estimated likelihood of an article containing a specific topic, we create thirteen subsamples, each of which consists of articles that are most likely to represent one particular topic. Specifically, for each topic $k = \{1, 2, \cdots, K\}$, we rank articles by each article i's probability of representing topic k, i.e., p_{ik}, and select articles from the top quartile.[29] Since LDA allows each document (an article in our case) to contain

[28] If the number of topics is too low, the lawsuit of Huawei executive Meng Wanzhou, a longstanding and high-profile news subject, will be classified with human rights issues such as Xinjiang. In contrast, if the number of topics is too high, multiple topics could share a common theme. For example, if we raise the number of topics beyond $K = 14$, we obtain two or more topics related to the COVID-19 crisis that are difficult to distinguish from each other.

[29] We have also experimented with higher or lower thresholds such as the top 20% or 30%. All the

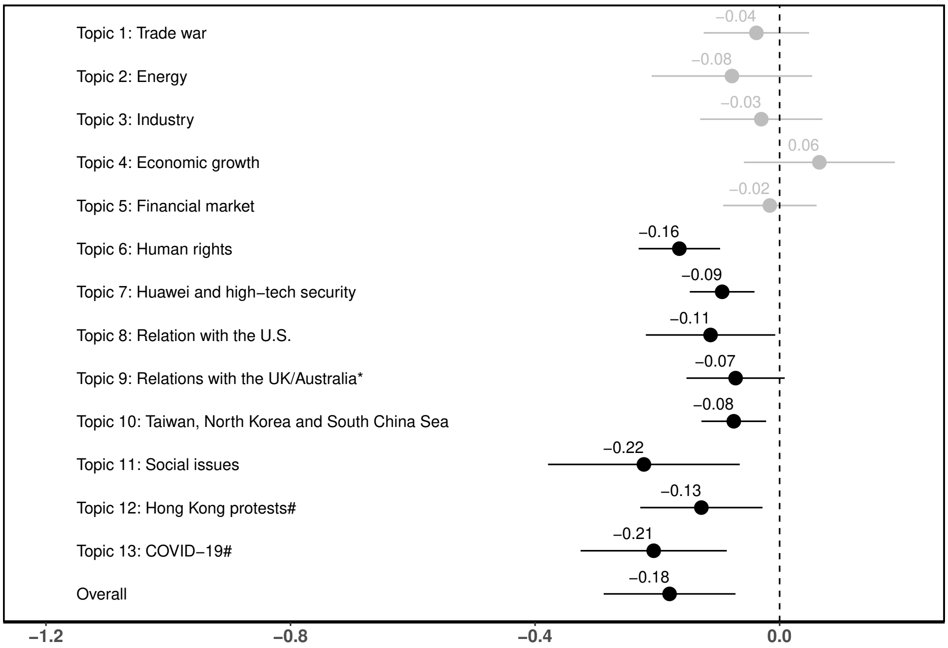


Figure 7. *Impacts at the Intensive Margin. The figure illustrates the difference-in-differences coefficient and the 95% confidence interval estimated for each news topic. There is no significant change in topics related to the Chinese economy, i.e., topics 1-5. However, the treated media became more negative toward China in politically sensitive topics, i.e., topics 6-11, in comparison to the control outlets. For topic 9, the $*$ notation indicates that the coefficient is significant at the 10% level. For topics 12 and 13, the # notation indicates that the coefficients are not interpretable.*

multiple topics, the subsamples are not mutually exclusive.

We estimate Equation (1) using each of the thirteen subsamples. In five of thirteen estimations, the coefficients of the interaction term $T \times Post$ are not significantly different from zero, while in the other eight estimations, the coefficients are all significantly negative. We group the former in Table 19 and the latter in Table 20, and both tables are relegated to Appendix F. We present the estimated coefficients and the 95% confidence intervals for all topics in Figure 7 for ease of comparison.

On topics 1-5, the treated media outlets did not respond to the blockage differently from the control group. These five topics are all related to the Chinese economy and traditionally considered within the redline of Chinese censors. The trade war topic, as discussed in section 2.2, became a sensitive issue only after the sudden upending of trade negotiations that heralded the crackdowns. The consistently insignificant blockage effects suggest that the media did not intentionally manage the tone on the topics within the red lines before the crackdown, and therefore did not have to adjust their coverage afterwards.

In contrast, the result for the topic of human rights — a topic constantly agitat-

results were robust and similar.

ing the Chinese government (topic 6, presented in column (1) of Table 20), shows that the blockage increased the magnitude of negativity of the media's tone by 0.164. This effect is significant at the 1% level. Similar patterns are observed for Huawei and high-tech security (topic 7), relations with the US (topic 8), relations with the UK and Australia (topic 9), Taiwan, North Korea and the South China Sea (topic 10), and social issues (topic 11). These topics are more political and typically more sensitive than economic topics. Our findings suggest that the media, after being kicked out, increased the negativity of their tone on topics likely hit a nerve with the Chinese government.

As the COVID-19 crisis and Hong Kong anti-extradition bill protests occurred after the crackdown, the results of the DID models for the subsamples focused on the relevant topics (topics 12 and 13) are not interpretable. While the models are still technically estimable, LDA may assign high probabilities of being related to these topics to some news articles published before the two events actually happened.[30] As shown in section 5.2, the main finding is not driven by the coverage of those two topics.

Of interest to us is not only whether the treated media adjusted their tone after the blockage but also how they did so. Word choice is important to the reader's formation of a perception of the news content. For example, China could be referred to as either the largest developing country or a communist regime, leaving distinct impressions on readers. News journalists and editors have a lot of room to adjust the wording of their articles to be friendly toward the Chinese regime or critical of it. We observe significant changes in wording: the treated news media would use aggressive phrases such as "human rights abuse" "genocide" or "re-education camps" more often after they were blocked, relative to the control media. Such phrases and the related discussions are more negative in tone than other words in similar topics and drive down the overall tone of news articles containing them. A more systematic investigation into the effect of blockage on word choice is relegated to Appendix G.[31]

6.2. Extensive Margin: Reporting Frequency across Topics

Next, we switch the focus to another important dimension of the reporting strategy and investigate how news outlets adjusted the coverage frequency of each topic after being blocked. To explore the extensive margin, we need to assign news articles in

[30]For example, news articles about the annual July protest in Hong Kong in 2018 are given high probabilities of being related to topic 12, and news articles about epidemic outbreaks in 2019 or earlier, which are unrelated to COVID-19, are given high probabilities of covering COVID-19-related topics. See "Pneumonic Plague Is Diagnosed in China"(November 13, 2019, The New York Times).

[31]Additionally, we study whether the tone changes that we identify arise mainly from changes in how news journalists or editors present facts or how they interpret and analyze facts. We show that media outlets are more likely to adjust the content of news analysis rather than twist the facts, assuming that they compromise their reporting. This is discussed in Appendix G.

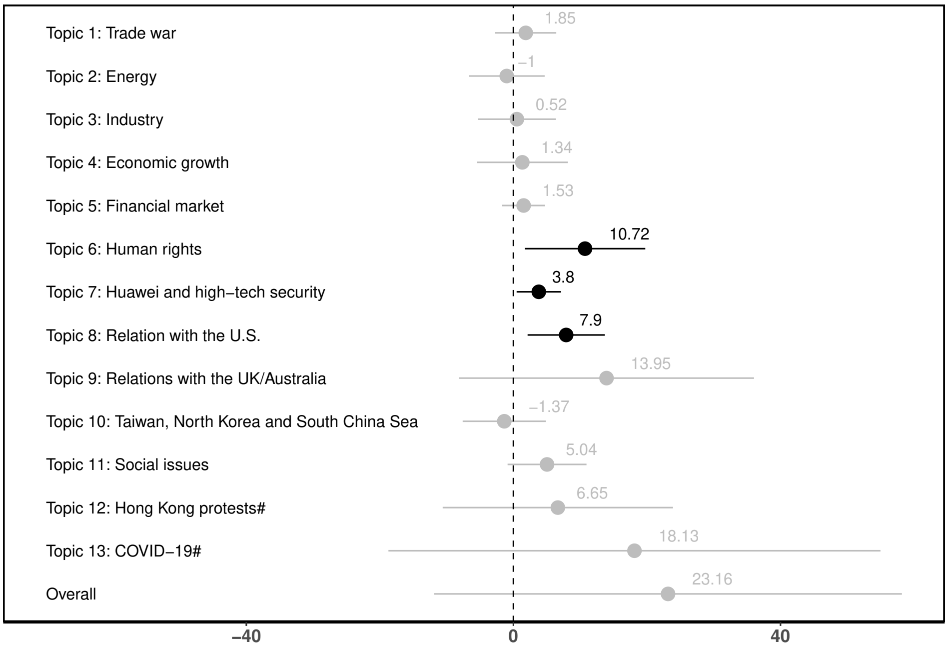


Figure 8. *Impacts at the Extensive Margin. The figure illustrates coefficients and the 95% confidence intervals estimated with the difference-in-differences model and by using the monthly number of articles by each news outlet in each news topic as dependent variables. There is no significant change in topics related to the Chinese economy, i.e., topics 1-5. However, there are significant changes in topics such as human rights, Huawei and high-tech security and relations with the US (i.e., topics 6-8): the treated outlets published more news articles on these topics after the blockage than did the control outlets. For topics 12 and 13, the # notation indicates that the coefficients are not interpretable.*

our sample to the thirteen news topics. To this end, we construct a dummy variable A_{ik} and assign the value of 1 to article i if article i's probability of representing topic k (i.e., p_{ik}) is in the top quartile among all articles (to be consistent with section 6.1), and set it to 0 otherwise. Then, we can sum the number of articles for each topic over each month for each media outlet. A summary of statistics is relegated to Table 21 in Appendix F.

To examine the changes in the monthly number of articles for each topic, we estimate a specification similar to Equation (1) but at the month-outlet level and with only month and outlet fixed effects as controls. We present results for topics 1-5 in Table 22 and those for topics 6-13 in Table 23; both are relegated to Appendix F. We plot the estimated DID coefficients and the 95% confidence intervals for all topics in Figure 8 for ease of comparison.

Our results suggest that the treated outlets, in comparison to the change in the control outlets, published 10 more articles per month per outlet on the topic of human rights, 4 more on Huawei and high-tech security and 8 more on relations with the US after the blockage, and that these effects are statistically significant. For topics related to the Chinese economy, i.e., topics 1-5, the difference between the treated and control groups did not change significantly after the blockage. These results indicate that the treated media increased their frequency of coverage of these rather sensitive topics, but not that of non-sensitive topics. In addition, we do not observe a significant rise in the treated media's overall frequency of reporting on China-related issues.

Two messages emerge from the exercises in both sections 6.1 and 6.2. First, the contrast across various subsamples is consistent with our conjecture that the media intentionally toned down their negativity toward China before the blockage, under the premise that the Chinese government was less tolerant of critical coverage of political issues such as human rights than of that of economics issues. Second, consistently with this conjecture, the media also suppressed the critical coverage of China by reducing the quantity of the news content on sensitive topics. In summary, our findings reveal one front along which the media compromise news reporting: they treat sensitive issues with caution, making their coverage less negative and covering them less often.

7. Interpretations

The interpretation of our findings thus far is that prior to the loss of access, news outlets optimized and managed their reporting styles by trading off their influence and profit at home and abroad, in the short run and the long run, while recognizing that Chinese censors might retaliate if the media crossed red lines. Once access was lost, media outlets had fewer constraints on choosing how and what to report.

Nevertheless, several alternative mechanisms are also plausible and equally interesting. For instance, the change in the reporting strategy might result from changes in editorial staff. Note that this mechanism does not apply to this setting because in most of our data period, journalists could still work in China and report on the crackdown themselves after their outlets were blocked (as discussed in section 2.3). There are a few other alternative mechanisms that demand close examination, and we proceed to consider them in this section.

7.1. Unleashing Grievances?

It is possible that victims of the crackdown were antagonized by the loss of influence or potential growth and hence adopted a more negative tone toward China to retaliate or express their grievance. Implicitly, this grievance interpretation assumes that the media did not intentionally tone down negativity toward China prior to the blockage but became harsher afterwards.

This grievance interpretation, albeit intuitive, is not well supported by the data. First, such a sense of grievance may be a likely reaction of the news production staff in the short run, but not very likely to be sustained in the long run. Gentzkow and

Table 6. Unleashing Grievances?: Default tone as outcome variable

	Triple Diff with Chinese Website		Triple Diff with Baidu Index	
	(1)	(2)	(3)	(4)
T $\times$ Post	-0.154**	-0.174***	-0.451**	-0.477***
	(0.061)	(0.059)	(0.166)	(0.164)
T $\times$ Post $\times I$(Chinese websites)	0.100	0.134*		
	(0.080)	(0.074)		
T $\times$ Post $\times$ ln(Baidu index)			0.064	0.076*
			(0.039)	(0.038)
T $\times I$(Chinese websites)	-0.263*			
	(0.138)			
T $\times$ ln(Baidu index)			-0.326***	-0.231**
			(0.062)	(0.093)
Post $\times I$(Chinese websites)	0.128**	0.071		
	(0.052)	(0.057)		
Post $\times$ ln(Baidu index)			0.026	0.027
			(0.027)	(0.026)
T	0.237*		1.503***	
	(0.134)		(0.282)	
I(Chinese websites)	0.056			
	(0.069)			
Post	-0.334***		-0.355**	
	(0.050)		(0.139)	
ln(Baidu index)			-0.005	-0.001
			(0.020)	(0.032)
Press FE	No	Yes	No	Yes
Month FE	No	Yes	No	Yes
Panel FE	Yes	Yes	Yes	Yes
R-square	0.105	0.142	0.113	0.142
N	35689	35689	35689	35689

Note: Controls include the total word count and the total occurrences of the word "China" and "Chinese" in the article. Standard errors in parentheses clustered at the media outlet level; * p<0.1, ** p<0.05, *** p<0.01.

Shapiro (2010) point out that the media's own views play a much smaller role in media bias than does the drive for profit maximization. The resentful treatment of China-related news would eventually stop if it failed to lead to any commercial returns. Our event study in section 5.2 illustrates that the blockage's impact on news tone did not dwindle over time, suggesting that the blockage effect did not arise solely from a short-run tantrum of the media.

Second, the grievance interpretation implies that media outlets became harsher toward China because the blockage hurt their commercial interests. If our estimated blockage effect mainly arose through this mechanism, we would expect that media outlets with more prior investment or influence in the Chinese market would suffer more from the crackdown and hence respond more vehemently.

Due to the lack of systematic data on media outlets' investment in China, we first measure their exposure to the Chinese market using the presence of Chinese websites officially run by the respective outlets. Out of the eighteen outlets in our sample, five have had Chinese websites (or have their news articles translated to Chinese regularly), namely the Wall Street Journal, the New York Times, the Washington Post, Reuters, and the Guardian. Having a Chinese website is not only a clear sign of interest in and effort toward developing the Chinese market but also likely to correlate with other vested interests in China. Therefore, we examine whether the blockage effect differs between outlets with Chinese websites and those without. We define an indicator "Chinese websites" for an outlet with Chinese websites and include in Equation (1) the second- and third-degree interactions among the Chinese websites' indicators, T, and $Post$. The results with main effects and fixed effects are reported in columns (1) and (2) of Table 6, respectively.

Similarly to our baseline results, the coefficient of T $\times$ Post is approximately -0.18 and is significant at the 1% level, whereas the coefficient of the triple interaction term Chinese website$\times$ T $\times$ Post is positive and has a relatively smaller magnitude of 0.13. This contrast implies that outlets that put substantial effort into developing the Chinese market exhibited a much weaker response to the abrupt blockage. This finding contradicts the grievance interpretation. Instead, it is more consistent with the self-censorship interpretation. The relevant news outlets, with their tangled business interests in China, did not want to offend the Chinese government too much because it could hurt them in other areas.

Next, we examine how the media's responses to the blockage differed by their influence in China. We use the Baidu search index for the news outlets' names as a proxy for their influence in China. Insofar as the search index measures Chinese readers' interest, it can also be a proxy for potential market demand for coverage from those outlets. The blockage would have resulted in a larger loss of potential readership in China for media outlets with more prior searches. If the grievance interpretation holds, we would expect more frequently searched media outlets to exhibit a stronger response to the blockage. To test this conjecture, we include the logarithm of the Baidu search index, i.e., ln(Baidu index), as a term in Equation (1), as well as the second- and third-degree interactions among ln(Baidu index), T and $Post$. The results with main effects and fixed effects are reported in columns (3) and (4) of Table 6, respectively. While the media became extra harsh toward China in response to the blockage, media outlets with more influence or exposure in China showed milder responses than did those with less influence among Chinese readers. Similarly to the results from the heterogeneity analysis regarding the presence of Chinese websites, this finding does not support the interpretation that the media turned hostile toward China out of grievance.

7.2. Responding to a Changed Composition of Readers?

Another alternative conjecture is that, having lost the Chinese audience because of the blockage, the treated media outlets adjusted their news materials to the taste of American and British readers. While this seems plausible, it would be a stretch to argue that the Chinese readers did not want to read materials about human rights, high-tech security issues (such as the topic of Huawei) as well as China's relations with the US.

Furthermore, if the changed reader composition was the primary mechanism at work, we would expect to observe a stronger response from media outlets with more exposure to or influence in China because they lost more readers and experienced a greater change in readers' composition. Contradicting this prediction, as shown in the previous subsection (i.e., in Table 6), media outlets with more exposure to the Chinese market tended to respond more mildly to the blockage, which suggests that the changed composition mechanism, if it exists, is unlikely to be the solo driver of our findings.

Nevertheless, we examine this concern directly. It would be ideal to have precise measures of the readership composition of each news outlet in China as well as the UK and the US. If this alternative interpretation were the primary driving force of our results, we would observe that the tone of the treated media exhibited no additional response to the blockage, relative to the control media, once we control for the Chinese and non-Chinese readerships. In other words, the changed readership composition would have explained the changed tone of treated media.

While such measures of readership are unavailable, we can still construct proxies for readership using attention of readers to the media. We use both the Baidu search index and Google trend data to this end. To proxy the attention of Chinese readers to each media outlet, we use the monthly level of the Baidu search index for the name of each newspaper (as discussed in section 3.1, page 11). The UK and US readers pay attention to the media for a wide variety of reasons, and obtaining information about China-related issues is likely to be merely a small part. Therefore, we use the monthly Google search frequency of the refined search term "newspaper name + China" in the UK and US domains to proxy the degree to which readers rely on that particular newspaper to obtain information about China.[32] For example, the search intensity of "The Washington Post China" in the US likely represents how often readers in the US search for the Washington Post to learn about China during that month.

Figure 9 illustrates the average Google Trends Index by group. It indicates that (i)

[32]The Google Trends website offers domain-based search intensity data.

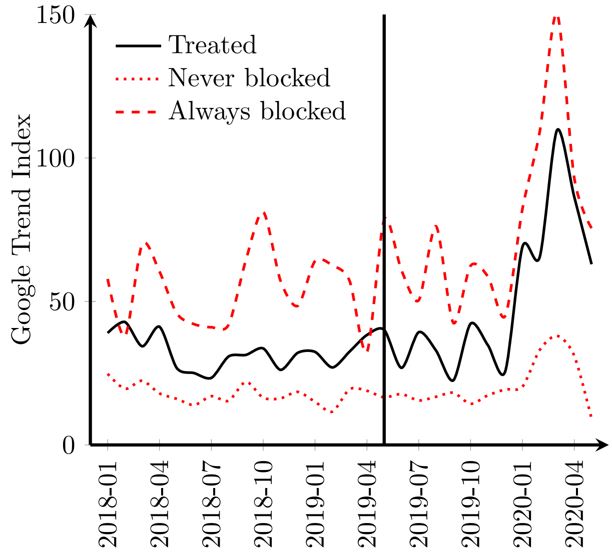


Figure 9. *Average Google Trends Index by group. The always-blocked media outlets are also more influential than others in the UK and the US in the domain of China-related issues. The media outlets newly blocked during the 2019 crackdown were searched for more often than the never-blocked outlets. The index for each group increased in February 2020, indicating that internet users in the UK and the US performed more searches for information about the COVID-19 pandemic (that originated in China) in the media.*

more influential media outlets are indeed searched for more often regarding China-related issues in the UK and the US, (ii) the search term "newspaper name + China" is informative as to readers' attention; e.g., readers in the UK and the US searched for "China" more often in the beginning of the COVID-19 pandemic, and (iii) it does not seem to be apparent that the reliance of readers' in the UK and the US on the treated media to obtain information about China increased or declined after such outlets were blocked.

Table 7 reports the results of estimating the DID model (1) with additional controls for the proxies for readerships. Columns (1) and (2) show the estimates for the DID model with main effects, controlling for Google Trends and the Baidu index, respectively. Column (3) shows the result of controlling for both indexes. In columns (4), (5) and (6), we report the corresponding results of estimating the DID model (i.e., Equation (1)) with fixed effects. In all specifications, the coefficients of T $\times$ Post are within the range from -0.175 to -0.2, are highly significant, and are very close to the baseline results reported in Table 4. This suggests that it is unlikely that the audience composition change is the only or major mechanism that drives the tone change of the treated media.

Table 7. *A Composition Change in Audience's Attention: Default tone as outcome variable*

	Main Effects			Fixed Effects		
	(1)	(2)	(3)	(4)	(5)	(6)
T × Post	-0.193***	-0.197**	-0.205***	-0.175***	-0.184***	-0.179***
	(0.061)	(0.069)	(0.064)	(0.049)	(0.047)	(0.044)
T	0.143	0.138	0.143			
	(0.101)	(0.097)	(0.096)			
Post	-0.239***	-0.238***	-0.238***			
	(0.038)	(0.039)	(0.040)			
ln(Google index)	-0.001		0.012	-0.017		-0.017
	(0.017)		(0.016)	(0.017)		(0.017)
ln(Baidu index)		-0.021	-0.034		-0.014	-0.013
		(0.027)	(0.022)		(0.045)	(0.046)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	No	No	No	Yes	Yes	Yes
Press FE	No	No	No	Yes	Yes	Yes
Panel FE	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.099	0.100	0.100	0.141	0.140	0.141
N	35,689	35,689	35,689	35,689	35,689	35,689

Notes: Controls include the total word count and the total occurrences of the word "China" and "Chinese" in the article. Standard errors in parentheses clustered at the media outlet level; * p<0.1, ** p<0.05, *** p<0.01.

7.3. Avoiding Offending the Chinese Readers?

Another alternative interpretation is that the media compromised their reporting before the blockage to avoid offending their Chinese audience (instead of the Chinese government) that might dislike negative news stories about China. This demand-side hypothesis is plausible and has been studied in the context of local news markets in the US (e.g., Gentzkow and Shapiro 2010). However, the premise of this interpretation has to be that Chinese audiences' news demand is aligned with what the Chinese government wants them to read. This is inconsistent with the government's enormous investment in censoring the internet (as discussed in section 2.1).

Next, we perform a few more empirical tests to complement our discussions. Our first strategy is to examine the coverage of news stories that are likely to be in demand among the Chinese public but are certainly disapproved of by the government. If coverage of these issues rose after the media lost their access, it would suggest that it is the government instead of the reader that is at the core of the media's calculation. A case in point concerns stories about scandals related to top leaders or their massive wealth; this topic was at the center of the newsroom drama at Bloomberg discussed in the introduction.

We construct a subsample of articles from our default sample that mention keywords related to corruption (variable construction is detailed in Appendix D). Presumably, these articles touch on corruption-related issues. We further create a dummy variable "scandal" that has the value of 1 if at least one sentence in the article mentions the wealth of or scandals related to top leaders (construction of relevant variables is again detailed in Appendix D). Using this subsample, we estimate Equation (1) with the scandal dummy variable being the outcome variable and report the results in column (1) of Table 8. Conditionally on reporting on corruption issues, the news outlets in the treatment group were 2% more likely to report on scandals related to top officials after the loss of access than those in the control group. Column (2) shows that the coverage of corruption-related issues by the treated outlets did not shrink in comparison to that by the control outlets. The results in columns (1) and (2) combined suggest that the treated outlets increased their coverage of top CCP leaders' scandals. This finding contradicts the explanation that the media censored themselves to please the Chinese readers.

The second strategy we rely on is to examine whether our baseline results are robust to including proxies for Chinese readers' news preferences as a control. If our estimated effect of the blockage arises mainly from the media no longer paying attention to Chinese readers' preferences, we would expect that a substantial part of the effect would be absorbed by the proxy for reader preference and its change.

The conjecture that the media avoided offending Chinese readers would not be relevant if the preferences of Chinese readers differed much from the voice of the Chinese government. Therefore, we take our test to an extreme by assuming that Chinese readers' preferences are close to those of the government. To this end, we proxy the preferences of Chinese readers using the news tone of the official mouthpiece of the government, namely China Daily. We construct the proxy "$Preference_n$," using the weekly average tone of China-related news articles published by China Daily in week n. Additionally, we construct a dummy variable "$Access_{jt}$," that has the value of 1 if press j was unblocked in month t and is 0 otherwise.

We first estimate Equation (1) by adding the variable $Preference_n$ as an additional control variable. Column (3) of Table 8 reports the result. The coefficient of the proxy $Preference_n$ is positive and significant, showing a correlation between the tone of China Daily and the media in our sample. Nevertheless, the estimated blockage effect (the coefficient of $T \times Post$) remains robust. We further include as a control the interaction of $Preference_n$ with $Access$. The coefficient of this new interaction term $Preference \times Access$ should capture the news media's response to the Chinese public's preferences when the media have access to the Chinese market. The results are reported in column (4) of Table 8. The estimated blockage effect, captured by the coefficient of the interaction $T \times Post$, is very similar to that in column (3) and

Table 8. Consumer Preference or Political Repercussions?

	Outcome Variables: Mention Scandal	Outcome Variables: Mention Corruption	Outcome Variables: Tone	Outcome Variables: Tone
	(1)	(2)	(3)	(4)
T × Post	0.021**	0.001	-0.180***	-0.201**
	(0.008)	(0.006)	(0.052)	(0.072)
Preference			0.127***	0.139***
			(0.017)	(0.013)
Preference × Access				-0.043
				(0.054)
Controls	Yes	Yes	Yes	Yes
Press FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Panel FE	Yes	Yes	Yes	Yes
R-Squared	0.052	0.037	0.144	0.144
N	1,346	35,689	33,114	33,114

Notes: Controls include the total word count and the total occurrences of the word "China" and "Chinese" in the article. Standard errors in parentheses clustered at the media outlet level; * p<0.1, ** p<0.05, *** p<0.01.

that of the baseline estimation (reported in column (2) of Table 4). The coefficient of $Preference \times Access$ is statistically insignificant. In summary, neither test supports the premise that the effect of the blockage was mainly driven by news outlets' concern with Chinese readers' preferences.

8. Concluding Remarks

It is not unlikely that free news media that enjoy the protection of the rule of law at home succumb to pressure from authoritarian regimes abroad. This phenomenon is new, partly because it has only been in recent decades that rising economic powers have happened to be undemocratic yet so economically intertwined or even coupled with democratic powers.

Our study advances the understanding of this issue by considering an episode in China that was sufficiently close to a quasi-natural experiment. Our findings suggest that, regarding the sensitive news topics, the news media may censor themselves and intentionally maintain a friendly tone in covering China, where they are keen on maintaining their presence. The news media choose to fine-tune their news products but do not interfere with opinion articles, for which they do not claim responsibility. In addition, when they are allowed to operate in China, the news media handle news issues that are sensitive in China with more caution than topics that are not sensitive. We provide suggestive evidence that news media compromise their reporting on China not because they are responding to demand from their Chinese audience but most likely because they fear retaliation from the authorities.

Authoritarian governments' possible manipulation of or interventions in news production have recently been an important issue in political discourse. Nevertheless, the discussions have centered mainly on the impact of direct interventions; e.g., foreign governments may wage disinformation campaigns or seek to control news outlets that target audiences in democratic countries. We discover a less apparent channel through which news production could be affected by foreign governments using economic leverage. This channel may pose no less of a threat to the backbone of democracy than outright interventions, given its concealed nature.

The mechanism underlying our findings is not unique to the news business. The Economist has recently observed that the global film industry is not free from meddling Chinese censors. Since China is becoming the world's largest cinema market by revenue, even overtaking America, Hollywood has geared its products to the Chinese market and, if necessary, altered films to please Chinese censors, including changing the versions for global audiences.[33] The other side of the coin is the case of Netflix, which has never been allowed to enter the Chinese market and therefore has had a free hand to commission documentaries about pro-democracy movements in Hong Kong, over which censors fret. In this paper, we did not deal with the potential impact of censorship stemming from foreign authoritarian regimes on citizens in democracies. We leave this important topic for future studies.

Our findings also beget new thinking on the censorship strategy of autocrats. Dealing with foreign entities—be it the New York Times or Hollywood—is tricky. Allowing such entities to have an influence at home unsurprisingly creates uneasiness for authoritarian regimes. However, autocrats who are eager to bolster their image overseas and who have economic power at their disposal lose the strings that they can pull behind the scenes when foreign entities are shut out entirely. The optimal degree of openness may require trading off their influence at home and abroad.

[33] "How Hollywood should deal with Chinese censors," and "Hollywood's Chinese conundrums," Aug 29, 2020, The Economist.

References

Baker, Scott R., Nicholas Bloom, and Steven J. Davis. 2016. "Measuring Economic Policy Uncertainty." *The Quarterly Journal of Economics* 131(4): 1593–1636.

Beattie, Graham, Ruben Durante, Brian Knight, and Ananya Sen. 2021. "Advertising Spending and Media Bias: Evidence from News Coverage of Car Safety Recalls." *Management Science* 67(2): 698–719.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-In-Differences Estimates?" *The Quarterly Journal of Economics* 119(1): 249–275.

Besley, Timothy, and Andrea Prat. 2006. "Handcuffs for the Grabbing Hand? Media Capture and Government Accountability." *American Economic Review* 96(3): 720–736.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *The Journal of Machine Learning Research* 3: 993–1022.

Catalinac, Amy. 2016. "From Pork to Policy: The Rise of Programmatic Campaigning in Japanese Elections." *The Journal of Politics* 78(1): 1–18.

Chen, Yuyu, and David Y. Yang. 2019. "The Impact of Media Censorship: 1984 or Brave New World?" *American Economic Review* 109(6): 2294–2332.

DellaVigna, Stefano, and Johannes Hermle. 2017. "Does Conflict of Interest Lead to Biased Coverage? Evidence from Movie Reviews. *Review of Economic Studies* 84(4): 1510–1550.

DellaVigna, Stephano, and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *The Quarterly Journal of Economics* 122(3): 1187–1234.

Di Tella, Rafael, and Ignacio Franceschelli. 2011. "Government Advertising and Media Coverage of Corruption Scandals. *American Economic Journal: Applied Economics* 3(4): 119–51.

Durante, Ruben, and Brian Knight. 2012. "Partisan Control, Media Bias, and Viewer Responses: Evidence from Berlusconi's Italy." *Journal of the European Economic Association* 10(3): 451–481.

Dyck, Alexander, and Luigi Zingales. 2003. "The Media and Asset Prices." Working Paper, Harvard Business School.

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. 2011. "Media and Political Persuasion: Evidence from Russia" *American Economic Review* 101(7): 3253–85.

Gentzkow, Matthew. 2006. "Television and Voter Turnout." *The Quarterly Journal of Economics* 121(3): 931–972.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as Data." *Journal of Economic Literature* 57(3): 535–74.

Gentzkow, Matthew, Nathan Petek, Jesse M. Shapiro, and Michael Sinkinson. 2015. "Do Newspapers Serve the State? Incumbent Party Influence on the US Press, 1869–1928." *Journal of the European Economic Association* 13(1): 29–61.

Gentzkow, Matthew, and Jesse M. Shapiro. 2006. "Media Bias and Reputation." *Journal of Political Economy* 114(2): 280–316.

Gentzkow, Matthew, and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence from US Daily Newspapers." *Econometrica* 78(1): 35–71.

Gentzkow, Matthew A. and Jesse M. Shapiro. 2004. "Media, Education and Anti-Americanism in the Muslim World." *Journal of Economic Perspectives* 18(3): 117–133.

Gerber, Alan S., Dean Karlan, and Daniel Bergan. 2009. "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions." *American Economic Journal: Applied Economics* 1(2): 35–52.

Germano, Fabrizio, and Martin Meier. 2013. "Concentration and Self-censorship in Commercial Media." *Journal of Public Economics* 97: 117–130.

Groseclose, Tim, and Jeffrey Milyo. 2005. "A Measure of Media Bias." *The Quarterly Journal of Economics* 120(4): 1191–1237.

Hansen, Stephen, Michael McMahon, and Andrea Prat. 2018. "Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach." *The Quarterly Journal of Economics* 133(2): 801–870.

Huang, Junming, Gavin Cook, and Yu Xie. 2021. "Do Mass Media Shape Public Opinion Toward China? Quantitative Evidence on New York Times With Deep Learning." *arXiv preprint arXiv:2012.07575.*

La Ferrara, Eliana, Alberto Chong, and Suzanne Duryea. 2012. "Soap Operas and Fertility: Evidence from Brazil." *American Economic Journal: Applied Economics* 4(4): 1–31.

Larcinese, Valentino, Riccardo Puglisi, and James M. Snyder . 2011. "Partisan Bias in Economic News: Evidence on the Agenda-setting Behavior of US Newspapers." *Journal of Public Economics* 95(9–10): 1178–1189.

MacKinnon, J. G. and M. D. Webb (2018). "The wild bootstrap for few (treated) clusters". *The Econometrics Journal 21*(2), 114–135.

McMillan, John, and Pablo Zoido. 2004. "How to Subvert Democracy: Montesinos in Peru." *Journal of Economic Perspectives* 18(4): 69–92.

Ozerturk, Saltuk. 2020. "Media Access, Bias and Public Opinion." Working paper, Southern Methodist University.

Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. 2014. "Glove: Global Vectors for Word Representation." In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543.

Prat, Andrea. 2018. "Media Power." *Journal of Political Economy* 126(4): 1747–1783.

Prat, Andrea, and David Strömberg. 2013. "The Political Economy of Mass Media." *Advances in Economics and Econometrics* 2: 135.

Qian, Nancy, and David Yanagizawa-Drott. 2017. "Government Distortion in Independently Owned Media: Evidence from US News Coverage of Human Rights." *Journal of the European Economic Association* 15(2): 463–499.

Rheault, Ludovic, Kaspar Beelen, Christopher Cochrane, and Graeme Hirst. 2016. "Measuring Emotion in Parliamentary Debates with Automated Textual Analysis." *PloS One* 11(12): e0168843.

Rosenbaum, Paul R. 2002. "Overt Bias in Observational Studies." In *Observational studies*, pp. 71–104. Springer, New York, NY.

Shapiro, Adam Hale, Moritz Sudhof, and Daniel J. Wilson. 2020. "Measuring News Sentiment." *Journal of Econometrics*.

Simonov, Andrey, and Justin Rao. 2022. "Demand for Online News under Government Control: Evidence from Russia." *Journal of Political Economy* 130(2).

Stanig, Piero. 2015. "Regulation of Speech and Media Coverage of Corruption: An Empirical Analysis of the Mexican Press. *American Journal of Political Science* 59(1): 175–193.

Strömberg, David. 2004. "Radio's Impact on Public Spending." *The Quarterly Journal of Economics* 119(1): 189–221.

Appendix

(Not intended for publication)

A. Tone Construction

The *GloVe* Algorithm

In this study, the tone of each article is an aggregation of each word in the text. To determine the tone of each word, we need to represent its meaning. One of the techniques of meaning representation is word embedding, i.e., representing a word by a dense and low-dimensional numerical vector in a meaningful manner. Given that some form of meaning is encoded in those vectors, semantic relations between words can be captured by the geometry of corresponding vectors. This work uses the algorithm of Global Vectors for Word Representation (*GloVe*), proposed by Pennington, Socher, and Manning (2014), to perform word embedding, which is one of the leading algorithms that excel in word analogy accuracy. *GloVe* is at least as efficient as the SKIM and CWOB methods. The algorithm is widely used and has been cited by more than 19,000 scientific articles so far.

First, it is essential for the *GloVe* algorithm to build the word-word co-occurrence matrix X, inside which each entry X_{ij} represents the number of times word j occurs in the context of word i, where context is defined as a window centered around the focus word. Therefore, the probability that word j appears in the context of word i is constructed by:

$$P_{ij} = \frac{X_{ij}}{X_i},$$

where X_i is the number of times any word appears in the context of word i.

Second, two features distinguish the *GloVe* method from others. (i) It utilizes the "co-occurrence probabilities ratios" rather than the raw probabilities. Pennington, Socher, and Manning (2014) show that the co-occurrence ratios gather more information and better capture the relationship between words. (ii) An efficient and workable function F is proposed to predict those ratios– such that

$$F\left(w_i, w_j, \tilde{w}_k\right) = \frac{P_{ik}}{P_{jk}}, \tag{5}$$

where w_i and w_j are two word vectors and $\tilde{w}_k$ is a context word vector.

One leading and frequently cited example that the authors use to illustrate this insights is as follows: "ice co-occurs more frequently with solid than it does with gas, whereas steam co-occurs more frequently with gas than it does with solid. Both words co-occur with their shared property water frequently, and both co-occur with the unrelated word fashion infrequently. Only in the ratio of probabilities does noise from non-discriminative words like water and fashion cancel out, so that large values (much greater than 1) correlate well with properties specific to ice, and small values (much less than 1) correlate well with properties specific of steam. In this way, the ratio of probabilities encodes some crude form of meaning associated with the abstract concept of thermodynamic phase (The *GloVe* official site)."

Third, equation (5) associates word vectors on the left-hand side with text statistics (i.e., those co-occurrence probabilities ratios) on the right hand side. That is, while those word vectors are to be learned, the probability ratios are observable empirically. A cost/objective function is defined to capture the differences between them. The *GloVe* algorithm minimizes this objective function by learning meaningful word vectors representations.

The News Corpus and Training

We need to feed the *GloVe* algorithm with a sufficiently large corpus so that the training process can generate word embedding vectors for each word in the corpus in a meaningful way. We therefore built a corpus that includes 22 news outlets in total: Breitbart News, Chicago Tribune, China Daily, Daily Mail, HuffPost, Los Angeles Times, NBC News, Newsday, New York Post, Reuters, San Francisco Chronicle, Star Tribune, The Boston Globe, The Dallas Morning News, The Guardian, The New York Times, The Straits Times, The Sydney Morning Herald, The Telegraph, The Wall Street Journal, The Washington Post, and USA Today. Those media are either in our control group, or treatment group, or included for the purpose of validation. We scraped articles from their websites that mention key words, i.e., China, Chinese, Hong Kong, HongKonger (HongKongese), Russia, Russian, Iran or Iranian, at least once. That corpus consists of more than 1,010,000 articles and 791,997,864 tokens.

We use the source code (written in C) provided by the authors. Specifically, the context window is chosen to be 15 words (both to the left and to the right), and the default number of word vector dimensions is 300 (a standard choice in the literature). The output of this training process is a datafile that contains vectors, each of which represents a word in our corpus. We repeated the same training process by choosing word vector dimensions to be 100 and 500.

Tone Construction

To measure the positivity/negativity of each word, we follow the algorithm proposed by Rheault, Beelen, Cochrane, and Hirst (2016). The key idea is that a word that is closer to a group of independently validated positive words and further away from a group of independently validated negative words, tends to be more positive in sentiment.

To operationalize this insight, Rheault, Beelen, Cochrane, and Hirst (2016) selected 100 positive seed words and 100 negative words on condition that the seed words are required to be neither polysemants nor analogies. The authors offer a complete list of the seed words in the appendix of their paper (see Tables H and I that list positive and negative ones, respectively). We use the same set of words for seed and their vector representations are extracted from the result of the training process using our news corpus.

Next, the distances between words are constructed with cosine similarity of word vectors. The similarity between w_i and w_j is:

$$\frac{w_i w_j}{||w_i||||w_j||}$$

where $||w_i||$ is the norm of word vector w_i and the similarity is in a $[-1, 1]$ interval. Intuitively, completely irrelevant words give a similarity score close to 0; two closely located vectors w_i and w_j in the space lead to a similarity score close to 1; antonym words generate a negative similarity.

Finally, to capture the net distance from the two sets of seed words, the emotion score of each word in our corpus is calculated as follows:

$$s_i = \sum_{p \in P} \frac{w_i w_p}{||w_i||||w_p||} - \sum_{q \in Q} \frac{w_i w_q}{||w_i||||w_q||},$$

where P is the 100 positive seed words set and Q is the 100 negative seed words set. A positive score s_i indicates that w_i is closer to positive seed words in the vector space than to the negative ones.

Using this approach, we can assign a score to every word in our corpus of news articles. Therefore, we built an emotional word lexicon with approximately 400,000 words, which have been used at least 5 times in the corpus. Its distribution is close to the normal but slightly negatively skewed with a mean value of -0.26 and a standard deviation of 2.95. Figure 10 illustrates the distribution of the emotion scores of words.

In our study, the emotion score (or the extent of positivity/negativity) of each news article is an aggregate of words in its text. To generate the scores, the standard pre-processing procedures are routinely followed: We first obtain the stop words consisting of English stop words in nltk package along with punctuation marks and names. For each text, we eliminate the stop words and convert all capital letters to lower case letters, etc. In general, by utilizing the word lexicon, we calculate the article level

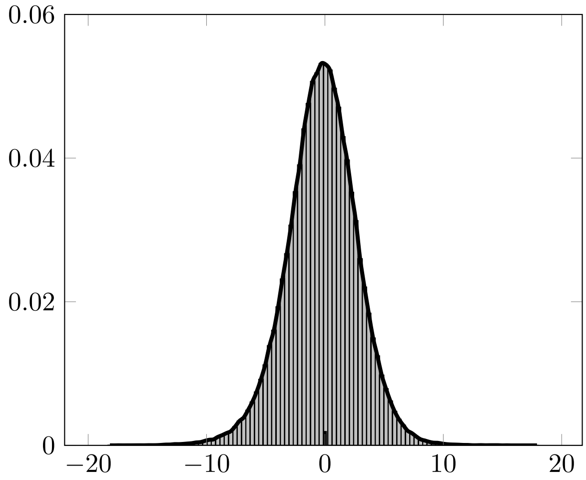


Figure 10. *Distribution of tone scores*

emotional score by following the procedure below:

a. For each text, generate the sentences in the text and split those to obtain word list. Note that we do not drop duplicates words.

b. For each word in the word list, find the corresponding score in the word lexicon and add it to the text score.

c. On condition that a word has a internal negation right before it, such as "not satisfying", we assign the opposite emotion value of this word's to this phrase.

d. The score of text is the sum of word scores in the word list divided by number of words.

Three primary text scores are constructed by varying the word list in the texts. First, we construct word lists by using *all* the sentences in the texts. Second, we only include sentences that mention “China” or “Chinese.” Third, we only include words whose emotion scores are far enough from the mean score of the lexicon, representing words with strong emotions, i.e., words whose scores are beyond one (or two) standard deviation(s) around the mean word score.

Article Level Validation

We adopt two approaches to validate our measure of tones at the article-level. First, our measure in fact assigns a continuum emotion value to each word with consideration of their meanings in this particular corpus, while the dictionary approach assigns positive and negative labels to a subset of words without such a concern. One reasonable validation is that our measure, if more informative, should be at least correlated

Table 9. *Sentiment scores and negative and positive words: default tone as outcome variable*

	NRC Lexicon		LSD Lexicon	
	(1)	(2)	(3)	(4)
Frac. of Negative Words	-25.36***	-23.35***	-22.82***	-22.61***
	(2.86)	(2.80)	(0.48)	(0.71)
Frac. of Positive Words	5.812***	6.052***	8.294***	8.090***
	(0.51)	(0.63)	(1.14)	(1.54)
Log Word Count	0.0854**	0.0817**	0.110**	0.089*
	(0.031)	(0.038)	(0.046)	(0.047)
Press FE	No	Yes	No	Yes
Month FE	No	Yes	No	Yes
Panel FE	No	Yes	No	Yes
R-square	0.3265	0.4010	0.4552	0.5253
N	35689	35689	35689	35689

Notes: Standard errors in parentheses clustered at the media outlet level; * p<0.1, ** p<0.05, *** p<0.01.

with those utilizing the traditional dictionary approaches, which are coarser. To test this idea, we construct the factions of positive and negative words in each article, by using two dictionaries, i.e., NRC Word-Emotion Association Lexicon (NRC thereafter) and Lexicoder Sentiment Dictionary (version 2015, LSD thereafter), respectively.

At the article level, we regress the default tone scores of our measure on the fractions of negative and positive words, controlling the logarithm of total word count of each article. The results by using NRC and LSD lexicon are reported in column (1) and (3) of Table 9, respectively. It is evident that our tone scores tend to be higher when the faction of negative words is lower and the faction of positive words is higher, and the correlation is highly significant. We also add a set of fixed effects, including press, month and panel fixed effects, and report the corresponding results in column (2) and (4). The magnitude and significance of these correlations barely change.

Second, we utilize human input as additional validation. We randomly draw 100 articles from our sample, and then asked four trained assistants, all of whom are native English speakers, to independently evaluate tones of those articles, i.e., labelling them as "very very negative (-3)", "very negative (-2)""negative (-1)", "neutral (0)", "positive (1)", "very positive (2)"and "very very positive (3)". We take the average of the individual scores as the average human rating for each article. We plot corresponding tone scores that are computed according to our algorithm against human ratings, as well as the fitted regression line in Figure 11. The estimated slope is 0.21 and it is highly significant, i.e., *p*-value is 0.005. There is a clear pattern whereby the computer

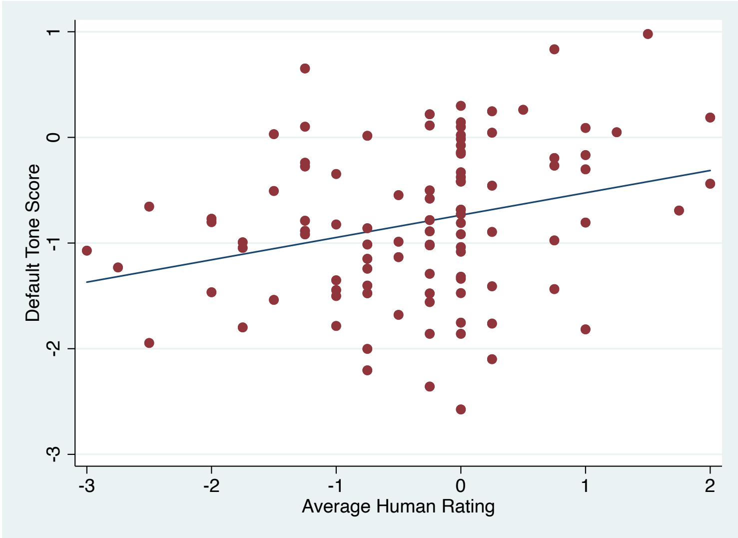


Figure 11. *Tone Score and Human Rating. The vertical axis is tone scores given by our algorithm and the horizontal axis shows average ratings of the human assistants.*

algorithm and human rating largely agree on the underlying tones of the articles.

To present a more concrete impression of the results of the algorithm that we use to compute tones, we select three articles from the New York Times in our sample, which were rated as relatively neutral, very negative and very positive by our algorithm. Mindful of the fact that the median tone score of the New York Times articles in our main sample is -0.70; the most negative -2.3, and the most positive 2.0. Below are three corresponding examples from the section of Asia-Pacific of the New York Times. We only show sample sentences that mentioned China or Chinese.

An article with an around-median score is "Trump Embraces Foreign Aid to Counter China's Global Influence (2018-10-14, score: -0.21)." Samples of sentences that mention "China or Chinese" are listed below:

> Mr. Trump seems to be learning that the projections of military power alone will not be enough to compete with *China*, he said.
>
> So much of our foreign policy now is focused on trying to check *China*, especially their nefarious activities.
>
> The key to its success, development officials said, is to create a new system that will carefully vet investments for maximum economic and political impact – and to ensure that projects don't fail as a result of corruption and mismanagement, a problem that has plagued *China*'s investments in Malaysia and elsewhere.
>
> A bigger question is whether it will do anything to reduce *China*'s global influence.

An article with very negative tone score is "Pneumonic Plague Is Diagnosed in China (2019-11-13, score: -2.28)." Samples of sentences that mention "China or Chinese" are listed below:

> On Tuesday, *Chinese* censors instructed online news aggregators in *China* to "block and control" online discussion related to news about the plague, according to a directive seen by The New York Times.
>
> Skeptical *Chinese* internet users have charged the government with being slow to disclose news about the disease, which is transmitted between humans and kills even faster than the more-common bubonic form.
>
> *China* has a history of covering up and being slow to announce infectious outbreaks, prompting many people to call for transparency this time.
>
> According to *China*'s health commission, six people have died in the country from the plague since 2014.

An article with very positive tone score is "Theater Director Returns to China With 'Liberating and Cool' Vision (2018-7-27, score: 1.58)." Samples of sentences that mention "China or Chinese" are listed below:

> In the way Chen Shi-Zheng imagines his theatrical adaptation of "The Orphan of Zhao," the production will bring out all the elements of the story that have appealed to *Chinese* audiences through the centuries, like the timeless themes of revenge and self-sacrifice.
>
> Over a recent dinner in New Haven, Mr. Chen and Audrey Li, his wife and business partner, talked with excitement about the chance for him to create a work for a *Chinese* audience again, playing the role of a cultural bridge as relations between the United States and *China* become more fraught over a variety of economic and security issues.
>
> After his formal arts education in *China*, he was invited to attend the Tisch School of the Arts at New York University as a graduate student, where he studied experimental theater from 1989 to 1991.

Table 10. *Excluding trade war and Tiananmen related articles*

	Samples Excluding Articles that Mention:			
	Trade War	Trade	TAM	Trade & TAM
	(1)	(2)	(3)	(4)
T × Post	-0.203***	-0.172***	-0.178***	-0.171***
	(0.049)	(0.046)	(0.053)	(0.049)
[*WB p-value*]	[0.024]	[0.061]	[0.048]	[0.068]
{*RI p-value*}	{0.020}	{0.019}	{0.040}	{0.022}
Controls	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Press FE	Yes	Yes	Yes	Yes
Panel FE	Yes	Yes	Yes	Yes
R-Squared	0.183	0.242	0.140	0.243
N	28,427	20,111	34,886	19,625

Notes: Controls include the total word count and the total occurrences of the word "China" and "Chinese" in the article. Standard errors in parentheses clustered at the media outlet level; * p<0.1, ** p<0.05, *** p<0.01. P-values computed with wild bootstrap and randomization inference are reported in the square and curly braces respectively.

B. Additional Empirical Results and Discussions

Driven by Trade war and Tiananmen?

Do news articles that mention the trade war and/or Tian'anmen drive the identified results? To address this issue, we remove articles that *ever* mention "trade war" and reestimate Equation (1). The results are reported in column (1) of Table 10. The estimated coefficient for the interaction term is still significant at the 1% level, and its magnitude is slightly larger. However, we worry that the single keyword for the trade war does not purge relevant articles completely. Therefore, we reestimate Equation (1) with a sample in which we remove articles that ever mention "trade" so that the news content is orthogonal to the trade war and report the results in column (2) of Table 10. By doing so, we drop approximately 40% of the sample, the estimate concerned is still significant at the 1% level, and its magnitude changes only slightly.

We restrict our sample to articles mentioning none of the keywords related to "Tiananmen" and reestimate Equation (1). The result remains robust, and the magnitude does not change (column (3) of Table 10), indicating that coverage of the anniversary is unlikely to drive the results. We repeat the same exercise with a sample that drops articles ever mentioning either "Tiananmen" or "trade". Column (4) of Table 10 reports the results, which are very similar to those in column (3). This suggests that the potential interaction of the two issues is also unlikely to matter.

For each specification in Table 10, we further compute the WB-based and RI-based p-values of the estimated blockage effects. All the p-values are below or slightly above

Table 11. Chilling Effects?

	Outcome Variable:			Outcome Variable:	
	Tone	China	Non-Neutral	Tone	Tone
	Default Sample			Large Sample	Small Sample
	(1)	(2)	(3)	(4)	(5)
$T^{Pseudo} \times$ Post	-0.068	-0.068	-0.083	-0.073	-0.065
	(0.061)	(0.075)	(0.090)	(0.056)	(0.073)
Controls	Yes	Yes	Yes	Yes	Yes
Press FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Panel FE	Yes	Yes	Yes	Yes	Yes
R-Squared	0.144	0.124	0.135	0.145	0.152
N	22,420	22,223	22,420	26,838	18,256

Notes: Controls include the total word count and the total occurrences of the word "China" and "Chinese" in the article. Standard errors in parentheses clustered at the media outlet level; * p<0.1, ** p<0.05, *** p<0.01.

5%, further indicating the robustness of these results.

Testing for the Chilling Effect: Heterogenous Responses within the Control Group?

We relabel the always-blocked media as the control group and consider the never-blocked media as the pseudo treatment group. We compared the changes in the tone of these two groups after the blockage by estimating the Equation (1). The result is reported in Table 11. Following the format of Table 3, columns (1)-(3) of Table 11 show the results for three measures of tone scores: the benchmark score, the China-based score, and the non-neutral scores. Columns (4) and (5) present results estimated using the large and small samples (as defined in section 3.1). None of the coefficients on the interaction term $T_j \times Post$ is statistically significant, which contradicts the conjecture that there are heterogenous responses across the two groups, or a chilling effect. The lack of chilling effect is consistent with the motivation of this particular crackdown event: The blocked media were selected based on influence instead of their prior news tones. Ruling out this possibility further bolsters our confidence in the validity of the control group.

Driven by Post-crackdown Events?

Could the harsher tone have arisen because the treated outlets by nature were more responsive to prominent newsworthy events taking place after the blockage? This may have occurred if outlets in the treatment and control groups differ in unobservable characteristics. To address this concern, we first conduct a robustness check by

Table 12. *Identifications Issues: Default tone as outcome variable*

	Samples Excluding Articles that Mention:		
	HK	COVID	HK & COVID
	(1)	(2)	(3)
T × Post	-0.176***	-0.098***	-0.072***
	(0.041)	(0.031)	(0.019)
[*WB p-value*]	[0.029]	[0.069]	[0.022]
{*RI p-value*}	{0.026}	{0.082}	{0.052}
Controls	Yes	Yes	Yes
Press FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Panel FE	Yes	Yes	Yes
R-Squared	0.116	0.111	0.065
N	26,152	28,806	21,206

Notes: Controls include the total word count and the total occurrences of the word "China" and "Chinese" in the article. Standard errors in parentheses clustered at the media outlet level; * p<0.1, ** p<0.05, *** p<0.01. P-values computed with wild bootstrap and randomization inference are reported in the square and curly braces respectively.

removing articles covering the most salient issues after the blockage and reestimate Equation (1) with the remaining sample. We consider two such examples— the 2019 prodemocracy protests in Hong Kong and the COVID-19 pandemic. Columns (1) and (2) of Table 12 present the results estimated by dropping articles that *ever* mention Hong Kong- and COVID-19-related keywords, respectively (see Appendix D for details). The result remains statistically significant at the 1% level. We further perform the same robustness check with a sample in which articles that *ever* mention either Hong Kong- or COVID-19-related keywords are purged and report the results in column (3) of Table 12. The estimate is still significant at the 1% level, but its magnitude is much smaller. The small WB-based and RI-based p-values of the estimated blockage effects, shown respectively in the square and curly braces in each column of Table 12, provides reassuring evidence for the robustness of the results. In section 6.1, we investigate how various news topics impact the estimate in the full sample with the aid of topic modeling techniques.

Robustness Tests: Measurements and Samples

To examine whether the results are robust to the measure of tone, we reestimate Equation (1) with alternative measures discussed in section 3.2. Columns (1) and (2) of Table 13 report the results using the China-based scores and the nonneutral scores, respectively. Consistent with the baseline results, losing access renders the tone of news

Table 13. Robustness: Alternative measures and samples

	Outcome Variable: China	Outcome Variable: Non-neutral	Outcome Variable: Wiki	Outcome Variable: Tone	Outcome Variable: Tone
	Default Sample			Large Sample	Small Sample
	(1)	(2)	(3)	(4)	(5)
T × Post	-0.157**	-0.268***	-0.145**	-0.169***	-0.168***
	(0.061)	(0.083)	(0.059)	(0.047)	(0.051)
Controls	Yes	Yes	Yes	Yes	Yes
Press FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Panel FE	Yes	Yes	Yes	Yes	Yes
R-Squared	0.118	0.132	0.152	0.141	0.144
N	35,380	35,689	35,689	42,190	29,739

Notes: Controls include the total word count and the total occurrences of the word "China" and "Chinese" in the article. Standard errors in parentheses clustered at the media outlet level; * p<0.1, ** p<0.05, *** p<0.01.

articles more negative. The estimated blockage effects on the China-based scores and the nonneutral scores are -0.157 and -0.268, respectively, corresponding to approximately 0.18 and 0.22 standard deviations of these two measures. The result is less significant by using China-based scores. It is expected, since the construction of this measure only involves a much smaller fraction of text in each article. Furthermore, we use the Wikipedia-based tone scores to cross-check our estimates, and the results remain robust (column (3) of Table 13). The estimated effect on the Wikipedia-based tone scores is -0.145, approximately 0.13 standard deviations, which is smaller and less significant (at the 5% level) than the effect on other measures derived from our own news corpus. Given that our word embedding approach is context based and corpus specific, using word vectors generated from other corpora inevitably introduces noise and measurement errors that bias the estimate toward zero and enlarge the standard errors.

Next, we test whether our results are robust to the choice of sample. We use two alternative samples, i.e., the large sample, which uses looser criteria and includes more articles than the default sample, and the small sample, which uses more stringent criteria and includes fewer articles (discussed in section 3.1). The results are reported in columns (4) and (5) of Table 13, respectively. The estimates are close to those estimated using the default news sample (column (2) of Table 4), suggesting that our results are robust to sample choices.

We trained the *GloVe* algorithm with our corpus, created a vector space of 300 dimensions and assigned each word in the corpus a corresponding vector. One may wonder whether our result is robust to the choice of dimension space. To address this concern, we re-train the algorithm twice and selected a vector space of 100 and 500 di-

Table 14. Robustness: Alternative dimensionality choices for GloVe

	100 Dimensions; Outcome Variables: Tone	China	Non-Neutral	500 Dimensions; Outcome Variables: Tone	China	Non-Neutral
	(1)	(2)	(3)	(4)	(5)	(6)
T $\times$ Post	-0.291***	-0.255**	-0.283***	-0.157***	-0.136**	-0.231***
	(0.089)	(0.106)	(0.087)	(0.043)	(0.051)	(0.066)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Press FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Panel FE	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.146	0.115	0.149	0.168	0.139	0.161
N	35,689	35,380	35,689	35,689	35,380	35,689

Notes: Controls include the total word count and the total occurrences of the word "China" and "Chinese" in the article. Standard errors in parentheses clustered at the media outlet level; * p<0.1, ** p<0.05, *** p<0.01.

mensions for each exercise. With the two sets of vectors, we can calculate the emotion score of each article and re-estimate our baseline equation.

Columns (1), (2) and (3) of Table 14 report the estimates respectively for default, China-based, and non-neutral tone scores constructed using 100-dimension vectors (counterparts of columns (2) of Table 4, (1) and (2) of Table 13 in the text). We repeated the entire exercise by using tone scores constructed using the 500-dimension vectors and reported the results in columns (4), (5) and (6). The dimensionality of vectors influences the measure of units. Therefore, it is not surprising that those estimates vary in size across different measurements. They turn out to be consistent in magnitude once we normalize the estimated impact with the standard deviation of each sample.

C. Excluding One Outlet

Table 15. *Excluding One Outlet*

Excluding:	β	S.E.	p-value
Breitbart News	-0.176	0.0626	0.0118
Chicago Tribune	-0.177	0.0525	0.00366
The Dallas Morning News	-0.181	0.0516	0.00269
Huffpost	-0.177	0.0532	0.00401
New York Post	-0.183	0.0507	0.00217
The New York Times	-0.185	0.0512	0.00213
Star Tribune	-0.179	0.0521	0.00321
The Boston Globe	-0.187	0.0494	0.00150
Daily Mail	-0.109	0.0502	0.0441
The Guardian	-0.201	0.0515	0.00116
Los Angeles Times	-0.177	0.0528	0.00383
Miami Herald	-0.182	0.0512	0.00249
NBC News	-0.187	0.0514	0.00208
Newsday	-0.180	0.0517	0.00286
Reuters	-0.132	0.0521	0.0212
San Francisco Chronicle	-0.185	0.0495	0.00162
USA Today	-0.181	0.0515	0.00270
The Washington Post	-0.208	0.0431	0.000158
The Wall Street Journal	-0.177	0.0523	0.00352

D. Key Words Construction

Freq. China & Chinese	The total number of occurrences of "China" and "Chinese" in one article.
Mention Tian'anmen	If the total number of occurrences of "Tian'anmen" is no-zero, it is equal to 1; otherwise it equals 0.
Mention HK	If the total number of occurrences of "Hong Kong", "HongKongese", "Hongkonger(s)" is no-zero, it is equal to 1; otherwise it equals 0.
Mention COVID	If the total number of occurrences of "covid", "coronavirus", "pandemic", "Wuhan virus", "China virus " or "Chinese virus " is non-zero, it is equal to 1; otherwise it equals 0.
Mention trade-war	If the total number of occurrences of "trade war " is non-zero, it equals 1; otherwise it equals 0.
Mention trade	If the total number of occurrences of "trade " is non-zero, it equals 1; otherwise it equals 0.
Mention corruption	If the total number of occurrences of the word "corruption", "corrupt", "corrupted", "corruptive", "corruptible", "corrupts "and "corrupting " is non-zero, it equals 1; otherwise it equals 0.
Mention scandal	If at least one of the words from the list of "asset", "wealth", "scandal " and at least one of the words from the list of "top chinese official", "top offical". "top leader", "paramount leader", "ccp leader", "chinese president", "party secretary " appear simultaneously in one sentence, it equals 1; otherwise, it equals 0.

E. Summary Statistics for the Russia and Iran Samples

Table 16. Summary of Statistics, Russia and Iran News Samples

	Russia			Iran		
	Treatment mean (sd)	Control mean (sd)	Diff mean (se)	Treatment mean (sd)	Control mean (sd)	Diff mean (se)
Default score	-1.05	-1.20	-0.15	-1.52	-1.69	-0.18
	(0.68)	(0.69)	(0.06)	(0.62)	(0.68)	(0.08)
Wordcount	1289.33	541.14	-748.19	1150.32	459.35	-690.96
	(2367.31)	(468.52)	(217.49)	(2294.02)	(352.38)	(296.08)
Freq. Russia & Russian	11.44	8.51	-2.94	0.80	0.45	-0.35
	(9.76)	(6.03)	(0.92)	(2.28)	(1.61)	(0.09)
Freq. Iran & Iranian	0.53	0.26	-0.27	15.39	10.70	-4.70
	(2.44)	(1.27)	(0.12)	(11.77)	(7.68)	(1.17)
N	3483	10056	13539	2388	7865	10253

Notes: The standard error in columns 3 and 6 are clustered at the press level.

F. Topic Modeling

To estimate the Latent Dirichlet Allocation (LDA) model, we pre-processed our news corpus by following standard practices. We converted every word in the corpus into lower case. We then cleaned the text by removing stop words that occur in the text as "noise", e.g., "a", "an" and "the" and removing punctuations; dashes within the word are preserved. Numbers, white space and URL are removed as well. We stem words in all texts, which allows us to reduce the size of document-term matrix. We only consider terms that occur at least five times in the corpus. As a result, the vocabulary size of the corpus becomes 40,466 and the LDA topic model is estimated with this preprocessed corpus.

For the LDA topic model, the number of topics K is of the most significant. In this paper, we choose $K = 13$ (justified in the main text) and fit the LDA topic model with Gibbs sampling. We follow the algorithm developed by Blei, Ng, and Jordan (2003) and implemented it in R with the *topic models* package. We tested the number of iterations for Gibbs sampling and found that the estimation results stabilized after 1,000 iterations. We also experimented with a higher number such as $K = 14$ or $K = 15$; the relevant results are rather similar.

We finally focused on two sets of important results from the estimation outputs: We obtained the most frequently used words in each topic and the distribution of each document over k topics. We interpreted the resulting topics by using the prior knowledge to associate them with the major and recurrent China-related events during the data period. Our results indicate that all the topics that emerge from our estimation are interpretable and intuitive, corresponding to identifiable news issues.

The topics uncovered by the estimated LDA model in terms of their highest-probability words are shown in Tables 17 and 18. We also illustrate topics used in the main text in the form of word clouds. See Figure 12. Words' probabilities of a given topic are in proportion to the size at which they are graphed.

Table 17. Top Word Lists

Topic 1 Trade War	Topic 2 Energy	Topic 3 Industry	Topic 4 Growth	Topic 5 Financial
trade	china	china	market	china
china	say	said	percent	said
trump	year	year	growth	compani
tariff	import	product	economi	bank
say	million	million	trade	invest
chines	oil	compani	year	year
deal	export	import	stock	billion
presid	price	sale	month	govern
state	product	industri	rate	will
unit	last	suppli	expect	firm
good	demand	price	global	busi
billion	produc	last	price	financi
import	suppli	will	econom	fund
talk	will	export	point	chines
beij	industri	produc	sinc	yuan
will	energi	oil	fell	market
administr	world	car	cut	develop
war	month	manufactur	last	new
american	percent	market	investor	also
negoti	crude	percent	index	polici

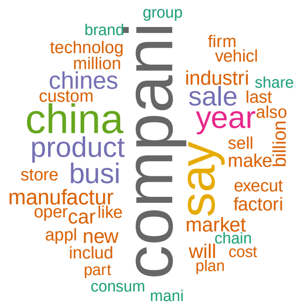
Table 18. *Top Word Lists (Cont'd)*

Topic 6 Human Rights	Topic 7 Huawei	Topic 8 Sino-US	Topic 9 UK/AUS	Topic 10 Taiwan	Topic 11 Social	Topic 12 HK Protest	Topic 13 Covid-19
china	say	presid	say	china	said	kong	coronavirus
chines	chines	trump	will	chines	year	hong	virus
parti	china	like	govern	taiwan	citi	protest	peopl
say	secur	say	minist	said	show	polic	case
communist	huawei	think	australia	militari	one	said	say
govern	offici	one	countri	countri	peopl	citi	health
beij	govern	time	new	south	famili	govern	outbreak
media	state	now	today	beij	chines	peopl	wuhan
peopl	report	can	australian	island	man	offic	test
report	foreign	american	also	region	day	demonstr	infect
right	investig	get	work	sea	two	mainland	china
offici	depart	want	servic	nation	video	bill	new
human	nation	just	now	will	children	lam	spread
countri	use	even	week	forc	build	law	report
xinjiang	compani	hous	britain	state	accord	beij	travel
social	technolog	make	may	airlin	home	extradit	hospit
polit	inform	state	told	intern	new	one	death
post	law	know	british	ship	provinc	use	day
state	case	peopl	prime	project	live	movement	patient
time	charg	way	time	road	told	fire	confirm

Word Clouds.

(a) Topic 1

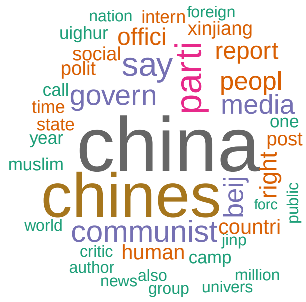
(b) Topic 2



(c) Topic 3

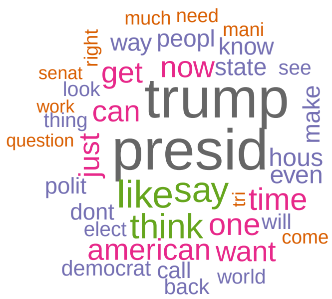
(d) Topic 4

(e) Topic 5



(f) Topic 6

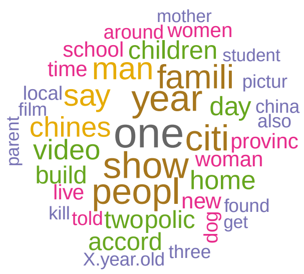
(g) Topic 7



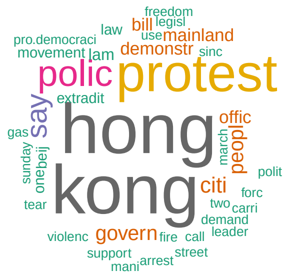
(h) Topic 8

(i) Topic 9

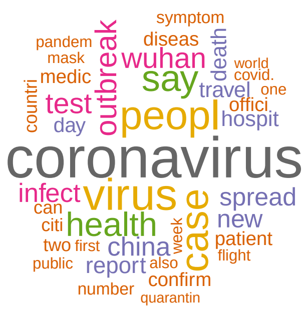
(j) Topic 10



(k) Topic 11



(l) Topic 12



(m) Topic 13

Figure 12. Word Clouds.

Table 19. Economic Topics: Intensive Margin

	Topic 1 Trade war (1)	Topic 2 Energy (2)	Topic 3 Industry (3)	Topic 4 Growth (4)	Topic 5 Financial (5)
T × Post	-0.038	-0.078	-0.030	0.065	-0.016
	(0.044)	(0.067)	(0.051)	(0.063)	(0.039)
Controls	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Press FE	Yes	Yes	Yes	Yes	Yes
Panel FE	Yes	Yes	Yes	Yes	Yes
R-Squared	0.089	0.100	0.101	0.121	0.098
N	8,924	8,921	8,923	8,918	8,925

Notes: Controls include the total word count and the total occurrences of the word "China" and "Chinese" in the article. Standard errors in parentheses clustered at the media outlet level; * p<0.1, ** p<0.05, *** p<0.01.

Table 20. Politically Sensitive Topics: Intensive Margin

	Topic 6 Human rights (1)	Topic 7 Huawei (2)	Topic 8 Sino-US (3)	Topic 9 UK/AUS (4)	Topic 10 Taiwan (5)	Topic 11 Social (6)	Topic 12 HK (7)	Topic 13 COVID (8)
T × Post	-0.164***	-0.094***	-0.113**	-0.072*	-0.075**	-0.222**	-0.128**	-0.206***
	(0.034)	(0.027)	(0.054)	(0.041)	(0.027)	(0.080)	(0.051)	(0.061)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Press FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.178	0.148	0.167	0.140	0.135	0.176	0.172	0.146
N	8,920	8,924	8,924	8,922	8,923	8,922	8,922	8,922

Notes: Controls include the total word count and the total occurrences of the word "China" and "Chinese" in the article. Standard errors in parentheses clustered at the media outlet level; * p<0.1, ** p<0.05, *** p<0.01.

Table 21. *Summary statistics: Monthly number of articles.*

		Treatment		Control	
Topic Number	Topic Name	mean	sd	mean	sd
1	Trade war	22.83	19.90	12.36	15.62
2	Energy	26.79	23.20	15.07	19.01
3	Industry	29.15	25.38	17.09	21.69
4	Growth	29.52	25.66	17.43	22.18
5	Financial	27.25	23.59	15.48	19.56
6	Human rights	27.17	23.51	15.41	19.48
7	Huawei	29.37	25.54	17.30	21.98
8	Sino-US	24.08	20.92	13.22	16.59
9	UK/AUS	33.28	28.42	20.80	27.40
10	Taiwan	34.97	29.66	22.46	30.11
11	Social Issues	29.91	25.96	17.80	22.77
12	HK protest	39.05	33.05	26.60	38.27
13	Covid-19	34.64	29.38	22.14	29.50

Table 22. Economic Topics: Extensive Margin

	Topic 1 Trade war (1)	Topic 2 Energy (2)	Topic 3 Industry (3)	Topic 4 Growth (4)	Topic 5 Financial (5)
T × Post	1.848	-1.001	0.518	1.336	1.532
	(2.332)	(2.897)	(2.985)	(3.476)	(1.638)
Month FE	Yes	Yes	Yes	Yes	Yes
Press FE	Yes	Yes	Yes	Yes	Yes
R-Squared	0.840	0.880	0.838	0.880	0.913
N	536	536	536	536	536

Notes: Standard errors in parentheses clustered at the media outlet level; * p<0.1, ** p<0.05, *** p<0.01.

Table 23. Politically Sensitive Topics: Extensive Margin

	Topic 6 Human rights (1)	Topic 7 Huawei (2)	Topic 8 Sino-US (3)	Topic 9 UK/AUS (4)	Topic 10 Taiwan (5)	Topic 11 Social (6)	Topic 12 HK (7)	Topic 13 COVID (8)
T × Post	10.724**	3.800**	7.904**	13.947	-1.366	5.036	6.654	18.127
	(4.606)	(1.694)	(2.949)	(11.268)	(3.181)	(3.016)	(8.799)	(18.798)
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Press FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.841	0.874	0.789	0.734	0.890	0.888	0.760	0.437
N	536	536	536	536	536	536	536	536

Notes: Standard errors in parentheses clustered at the media outlet level; * p<0.1, ** p<0.05, *** p<0.01.

Table 24. *Tone changes in subsamples of various lengths, default tone as outcome variable*

	Short	Medium	Long	Very Long
	(1)	(2)	(3)	(4)
T × Post	0.013	-0.138***	-0.225***	-0.190**
	(0.081)	(0.043)	(0.065)	(0.077)
Controls	Yes	Yes	Yes	Yes
Press FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Panel FE	Yes	Yes	Yes	Yes
R-Squared	0.154	0.133	0.160	0.195
N	8,907	8,878	8,889	8,920

Notes: Controls include the total word count and the total occurrences of the word "China" and "Chinese" in the article. Standard errors in parentheses clustered at the media outlet level; * p<0.1, ** p<0.05, *** p<0.01.

G. Analysis and Wording

News Analysis vs. Briefings

One question that we intend to explore in this section is whether the tone changes that we identify arise mainly from changes in the way that news journalists or editors present facts or the way that they interpret and analyze facts. It is challenging to separate facts from analysis in a given news report. Therefore, we turn to an indirect approach, which involves separating articles that are more likely to be news briefings from those that are more likely to be analytical and investigative reports. To disentangle the two types, we make use of the information on article length, under the assumption that the longer an article is, the more likely it is to be an investigation or analytical report and less likely to be a fact briefing piece. We then examine the pattern of tone changes for each type.

We divide our main sample into four quartiles based on the length of the articles, subsequently labeling them the short quartile, the medium quartile, the long quartile and the very long quartile. Columns (1)-(4) of Table 24 present the results from estimating Equation (1) using the four subsamples. The estimated effects of the blockage are statistically insignificant for the short quartile subsample and significant for the other three quartiles. Regarding the magnitude, the estimated effects for the long and very long quartiles of articles are much larger than those for the medium quartile. The results suggest that the tone changes caused by losing market access were likely to occur in news reports with analytical and investigative elements rather than news briefings focusing on facts.

We interpret this set of results as evidence that journalists and editors adopted

Table 25. Wording: Changes in fractions of positive and negative words used

	Outcome Variables:					
	No Exclusion		Ex 1 Std, Strong		Ex 2 Std, Strong	
	% Pos.	% Neg.	% Pos.	% Neg.	% Pos.	% Neg.
	(1)	(2)	(3)	(4)	(5)	(6)
T × Post	-0.018***	0.019***	-0.017***	0.019**	-0.013***	0.015***
	(0.005)	(0.006)	(0.005)	(0.007)	(0.003)	(0.005)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Press FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Panel FE	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	0.127	0.159	0.113	0.150	0.079	0.125
N	35,689	35,689	35,689	35,689	35,689	35,689

Notes: Controls include the total word count and the total occurrences of the word "China" and "Chinese" in the article. Standard errors in parentheses clustered at the media outlet level; * p<0.1, ** p<0.05, *** p<0.01.

a more negative tone while analyzing China-related news issues once they became less worried about offending Chinese censors. In other words, in compromising their reporting, media outlets are more likely to adjust the content of news analysis rather than twist the facts.

Wording Choices

News writers have a large room to adjust the wording of their articles, which could leave quite different impressions on readers in terms of author tone. For example, writers may refrain from using politically and emotionally charged phrases such as "massacre", which is very negative in tone, and replace it with "movement" or even "event" , which is less negative, or avoid mentioning an incident altogether. As the general news tone deteriorated after treated media were blocked, a follow-up question is whether journalists and editors adjusted their wording by reducing the usage of positive words or increasing the usage of negative words or both.

To investigate, we construct two measures of the composition of emotional words in each article, one representing the fraction of positive words (whose emotional value is above zero) used in the entire article and the other representing the fraction of negative words (whose emotional value is below zero) used. We estimate Equation (1) using the two fractions as outcome variables and present the results in columns (1) and (2) of Table 25, respectively. Both estimates are statistically significant, suggesting that the treated media outlets tended to adjust on both fronts after being blocked, using positive words less frequently and negative words more frequently than their counterparts in the control media outlets.

Does this effect remain if we count only words with strong emotions? We compute for each article the fraction of strong positive words (whose emotional value is half a standard deviation above the mean value of the lexicon) and strong negative words (whose emotional value is half a standard deviation below the mean). Columns (3) and (4) of Table 25 report the results from estimating Equation (1). Columns (5) and (6) reperform the exercise by resetting the threshold for defining strong positive and negative words to be one standard deviation above or below the mean value. All the results remain consistent and robust.