ARTICLE

Check for updates

Large-scale quantitative evidence of media impact on public opinion toward China

Junming Huang [1], Gavin G. Cook[1] & Yu Xie[1,2]✉

Do mass media influence people's opinions of other countries? Using BERT, a deep neural network-based natural language processing model, this study analyzes a large corpus of 267,907 China-related articles published by The New York Times since 1970. The output from The New York Times is then compared to a longitudinal data set constructed from 101 cross-sectional surveys of the American public's views on China, revealing that the reporting of The New York Times on China in one year explains 54% of the variance in American public opinion on China in the next. This result confirms hypothesized links between media and public opinion and helps shed light on how mass media can influence the public opinion of foreign countries.

[1] Princeton University, Princeton, NJ, USA. [2] Peking University, Beijing, China. ✉email: yuxie@princeton.edu

Introduction

America and China are the world's two largest economies, and they are currently locked in a tense rivalry. In a democratic system, public opinion shapes and constrains political action. How the American public views China thus affects relations between the two countries. Because few Americans have personally visited China, most Americans form their opinions of China and other foreign lands from media depictions. Our paper aims to explain how Americans form their attitudes on China with a case study of how The New York Times may shape public opinion. Our analysis is not causal, but it is informed by a causal understanding of how public opinion may flow from the media to the citizenry.

Scholars have adopted a number of wide-ranging and even contradictory approaches to explain the relationships between media and the American mind. One school of thought stresses that media exposure shapes public opinion (Baum and Potter, 2008; Iyengar and Kinder, 2010). Another set of approaches focuses on how the public might lead the media by analyzing how consumer demand shapes reporting. Newspapers may attract readers by biasing coverage of polarizing issues towards the ideological proclivities of their readership (Mullainathan and Shleifer, 2005), and with the advent of social media platforms such as Facebook and Twitter, traditional media are now more responsive to audience demand than ever before (Jacobs and Shapiro, 2011). On the other side of this equation, news consumers generally tend to seek out news sources with which they agree (Iyengar et al., 2008), and politically active individuals do so more proactively than the average person (Zaller, 1992).

Two other approaches address factors outside the media–public binary. The first, stresses the role of elites in opinion formation. While some, famously including Noam Chomsky, argue that news media are unwitting at best and at worst complicit "shills" of the American political establishment, political elites may affect public opinion directly by communicating with the public (Baum and Potter, 2008). Foreign elites may also influence American opinion because American reporters sometimes circumvent domestic sources and ask trusted foreign experts and officials for opinions (Hayes and Guardino, 2011). The second stresses how the macro-level phenomenon of public sentiment is shaped by micro-level and meso-level processes. An adult's opinions on various topics emerge from their personal values, many of which are set during and around adolescence from factors outside of the realm of individual control (Hatemi and McDermott, 2016). Social networks may also affect attitude formation (Kertzer and Zeitzoff, 2017).

In light of these contradictory interpretations, it is difficult to be sure whether the media shape the attitudes of consumers or, on the other hand, whether consumers shape media (Baum and Potter, 2008). Moreover, most of the theories summarized above are tested on relatively small slices of data. In order to offer an alternative, "big data"-based contribution to this ongoing debate, this study compares how the public views China and how the news media report on China with large-scale data. Our data set, which straddles 50 years of newspaper reporting and survey data, is uniquely large and includes more than a quarter-million articles from The New York Times.

Most extant survey data indicate that Americans do not seem to like China very much (Xie and Jin, 2021). Many Americans are reported to harbor doubts about China's record on human rights (Aldrich et al., 2015; Cao and Xu, 2015) and are anxious about China's burgeoning economic, military, and strategic power (Gries and Crowson, 2010; Yang and Liu, 2012). They also think that the Chinese political system fails to serve the needs of the Chinese people (Aldrich et al., 2015). Most Americans, however, recognize a difference between the Chinese state, the Chinese people, and Chinese culture, and they view the latter two more

favorably (Gries and Crowson, 2010). In Fiske's Stereotype Content Model (Fiske et al., 2002), which expresses common stereotypes as a combination of "competence" and "warmth", Asians belong to a set of "high-status, competitive out-groups" and rank high in competence but low in warmth (Lin et al., 2005).

The New York Times, which calls itself the "Newspaper of Record", is the most influential newspaper in the USA and possibly even in the Anglophonic world. It boasts 7.5 million subscribers (Business Wire, 2021), and while the paper's reach may be impressive, it is yet more significant that the readership of The New York Times represents an elite subset of the American public. Print subscribers to The New York Times have a median household income of $191,000, three times the median income of US households writ large (Rothbaum and Edwards, 2019). Despite the paper's haughty and sometimes condescending reporting, it "has had and still has immense social, political, and economic influence on American and the world" (Schwarz, 2012, p. 81). The New York Times may be a paper for America's elite, and it may be biased to reflect the tastes of its elite audience, but the paper's ideological slant does not affect our analyses as long as the its relevant biases are consistent over the time period covered by our analyses. Our analyses support the intuition of qualitative work on The Times (Schwarz, 2012) and show that these biases remain more or less constant for the decades in our sample. These analyses also illuminate some of the paper's more notable biases, including the paper's particular predilection for globalization.

The impact of social media on traditional media is not straightforward. While new media have certainly changed old media, neither has replaced the other. It is more accurate to say that old media have been integrated into new media and, in some ways, become a form of new media themselves. Twitter has accelerated the 2000s-era trends of information access that made it possible for news readers to find their own news and also enabled readers to interact with journalists (Jacobs and Shapiro, 2011), and the The New York Times seems to have made a significant commitment to the Twitter ecosystem. A quick glance at the follower count of The Times' official Twitter account shows that it is one of the most influential accounts on the site, with almost 50 million followers. For comparison, both current president Joe Biden and vice president Kamala Harris have around 10 million followers. Most New York Times reporters additionally have "verified" accounts on the platform, which means that individual reporters may be incentivized to maintain public-facing profiles more now than in the past.

The media consumption patterns that made new media possible have changed the way The New York Times interacts with its audience and how it extracts revenue. The New York Times boasts a grand total of 7.5 million subscribers, but only 800,000 of them subscribe to the print edition. The Times' digital subscription base has boomed since the election of Donald J. Trump, growing almost sixfold from a paltry 1.3 million in 2015 to a staggering 6.7 million in 2020 (Business Wire, 2021). The Times increasingly relies more on digital subscriptions and less on print subscriptions and ad sales for revenue (Lee, 2020). Ad revenue for most papers has been in sharp decline since the early 2000s (Jacobs and Shapiro, 2011), and this trend has only continued into the present. The New York Times now operates almost like a direct-to-consumer, subscription tech startup. New media have not replaced but have certainly changed old media. The full impact of these changes is beyond the scope of this paper, and we suggest it as an area for further research.

A small body of prior work has studied the The New York Times and how The New York Times reports on China. Blood and Phillips use autoregression methods on time series data to predict public opinion (Blood and Phillips, 1995). Wu et al. use a similar autoregression technique and find that public sentiment regarding the

Table 1 Accurately quantifying media sentiment.

	Article sentiment (%)	Paragraph assignment (%)	Paragraph sentiment (%)
Ideology	88.7 ± 3.9	94.1 ± 1.9	90.4 ± 7.4
Government Administration	62.7 ± 6.6	95.6 ± 0.8	78.1 ± 4.3
Democracy	87.7 ± 5.2	95.7 ± 1.1	77.5 ± 5.4
Economic development	81.2 ± 6.3	94.9 ± 1.1	79.9 ± 5.0
Marketization	91.3 ± 4.3	96.9 ± 1.7	74.9 ± 14.5
Welfare Well-being	82.0 ± 4.4	94.5 ± 2.7	73.2 ± 9.8
Globalization	77.4 ± 5.8	93.3 ± 1.5	80.7 ± 5.6
Culture	79.8 ± 4.8	90.0 ± 1.9	73.5 ± 5.4

Media article sentiment accuracy is measured by the percentage of correctly labeled articles (positive, negative, neutral/irrelevant) in each topic. Paragraph assignment accuracy is measured by AUC (area under ROC curve) of binary classifions on each paragraph (positive/negative vs. neutral/irrelevant). Paragraph sentiment accuracy is measured by AUC of binary classifions on each paragraph (positive vs. negative).

economy predicts economic performance and that people pay more attention to economic news during recessions (Wu et al., 2002). Peng finds that coverage of China in the paper has been consistently negative but increasingly frequent as China became an economic powerhouse (Peng, 2004). There is very little other scholarship that applies language processing methods to large corpora of articles from The New York Times or other leading papers. Atalay et al. is an exception that uses statistical techniques for parsing natural languages to analyze a corpus of newspaper articles from The New York Times, the Wall Street Journal, and other leading papers in order to investigate the increasing use of information technologies in newspaper classifieds (Atalay et al., 2018).

We explore the impact of The New York Times on its readers by examining the general relationship between The Times and public opinion. Though some might contend that only elites read NYT, we have adopted this research strategy for two reasons. If the views of NYT only impacted the nation's elite, the paper's views would still propagate to the general public through the elites themselves because elites can affect public opinion outside of media channels (Baum and Potter, 2008). Additionally, it is a widely held belief that NYT serves as a general barometer of an agenda-setting agent for American culture (Schwarz, 2012). Because of these two reasons, we interpolate the relationship between NYT and public opinion from the relationship between NYT and its readers, and we extrapolate that the views of NYT are broadly representative of American media.

Our paper aims to advance understanding of how Americans form their attitudes on China with a case study of how The New York Times may shape public opinion. We hypothesize that media coverage of foreign nations affects how Americans view the rest of the world. This reduced-form model deliberately simplifies the interactions between audience and media and sidesteps many active debates in political psychology and political communication. Analyzing a corpus of 267,907 articles on China from The New York Times, we quantify media sentiment with BERT, a state-of-the-art natural language processing model with deep neural networks, and segment sentiment into eight domain topics. We then use conventional statistical methods to link media sentiment to a longitudinal data set constructed from 101 cross-sectional surveys of the American public's views on China. We find strong correlations between how The New York Times reports on China in one year and the views of the public on China in the next. The correlations agree with our hypothesis and imply a strong connection between media sentiment and public opinion.

Methods

We quantify *media sentiment* with a natural language model on a large-scale corpus of 267,907 articles on China from The New York Times published between 1970 and 2019. To explore sentiment from this corpus in greater detail, we map every article to a sentiment category (positive, negative, or neutral) in eight topics: ideology, government and administration, democracy, economic development, marketization, welfare and well-being, globalization, and culture.

We do this with a three-stage modeling procedure. First, two human coders annotate 873 randomly selected articles with a total of 18,598 paragraphs expressing either positive, negative, or neutral sentiment in each topic. We treat irrelevant articles as neutral sentiments. Secondly, we fine-tune a natural language processing model Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2018) with the human-coded labels. The model uses a deep neural network with 12 layers. It accepts paragraphs (i.e., word sequences of no more than 128 words) as input and outputs a probability for each category. We end up with two binary classifiers for each topic for a grand total of 16 classifiers: an *assignment* classifier that determines whether a paragraph expresses sentiment in a given topic domain and a *sentiment* classifier that then distinguishes positive and negative sentiments in a paragraph classified as belonging to a given topic domain. Thirdly, we run the 16 trained classifiers on each paragraph in our corpus and assign category probabilities to every paragraph. We then use the probabilities of all the paragraphs in an article to determine the article's overall sentiment category (i.e., positive, negative, or neutral) in every topic.

As demonstrated in Table 1, the two classifiers are accurate at both the paragraph and article levels. The assignment classifier and the sentiment classifier reach classification accuracy of 89–96% and 73–90%, respectively, on paragraphs. The combined outcome of the classifiers, namely article sentiment, is accurate to 62–91% across the eight topics. For comparison, a random guess would reach an accuracy of 50% on each task (see Supplementary Information for details).

American public opinion towards China is a composite measure drawn from national surveys that ask respondents for their opinions on China. We collect 101 cross-sectional surveys from 1974 to 2019 that asked relevant questions about attitudes toward China and incorporate a probabilistic model to harmonize different survey series with different scales (e.g., 4 levels, 10 levels) into a single time series, capitalizing on "seaming" years in which different survey series overlapped (Wang et al., 2021). For every year, there is a single real value representing American sentiment on China relative to the level in 1974. Put another way, we use sentiment in 1974 as a baseline measure to normalize the rest of the time series. A positive value shows a more favorable attitude than that in 1974, and a negative value represents a less favorable attitude than that in 1974. Because of this, the trends in sentiment changes year-over-year are of interest, but the absolute values of sentiment in a given year are not. As shown in Fig. 1, public opinion towards China has varied greatly from 1974 to 2019.
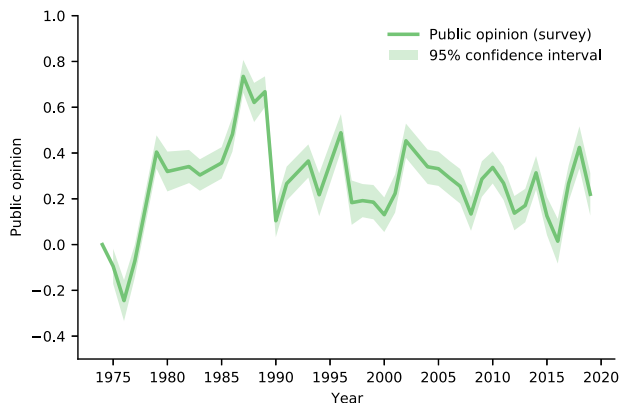
Fig. 1 Public opinion of Americans toward China. This time series is aggregated from 101 cross-sectional surveys from 1974 to 2019 that asked relevant questions about attitudes toward China with the year of 1974 as baseline. Years with attitudes above zero show a more favorable attitude than that in 1974, with a peak of 73% in 1987. Years with attitudes below zero show a less favorable attitude than that in 1974, with the lowest level of −24% in 1976. The time series is shown with a 95% confidence interval.
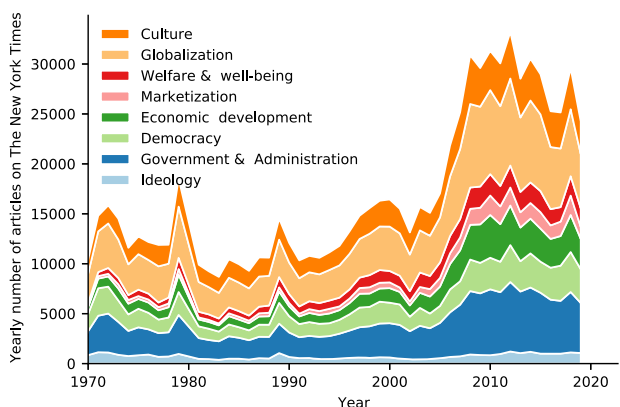
Fig. 2 Topic-specific yearly volume of The New York Times articles with sentiment on China from 1970 to 2019. In each year we report in each topic the number of positive and negative articles while ignoring neutral/irrelevant articles. The media have consistently high attention on reporting China government & administration, democracy, globalization, and culture. There are emerging interests on China's economics, marketization, and welfare and well-being since 1990s. Note that the sum of the stacks does not equal to the total volume of articles about China, because each article may express sentiment in none or multiple topics.

It steadily climbed from a low of −24% in 1976 to a high of 73% in 1987, and has fluctuated between 10% and 48% in the intervening 30 years.

Results

We begin with a demonstration of how the reporting of The New York Times on China changes over time, and we follow this with an analysis of how coverage of China might influence public opinion toward China.

Trend of media sentiment. The New York Times has maintained a steady interest in China over the years and has published at least 3,000 articles on China in every year of our corpus. Figure 2

displays the yearly volume of China-related articles from The New York Times on each of the eight topics since 1970. Articles on China increased sharply after 2000 and eventually reached a peak around 2010, almost doubling their volume from the 1970s. As the number of articles on China increased, the amount of attention paid to each of the eight topics diverged. Articles on government, democracy, globalization, and culture were consistently common while articles on ideology were consistently rare. In contrast, articles on China's economy, marketization, and welfare were rare before 1990 but became increasingly common after 2000. The timing of this uptick coincided neatly with worldwide recognition of China's precipitous economic ascent and specifically the beginnings of China's talks to join the World Trade Organization.

While the proportion of articles in each given topic change over time, the sentiment of articles in each topic is remarkably consistent. Ignoring neutral articles, Figure 3 illustrates the yearly fractions of positive and negative articles about each of the eight topics. We find four topics (economics, globalization, culture, and marketization) are almost always covered positively while reporting on the other four topics (ideology, government & administration, democracy, and welfare & well-being) is overwhelmingly negative.

The NYT views China's globalization in a very positive light. Almost 100% of the articles mentioning this topic are positive for all of the years in our sample. This reveals that The New York Times welcomes China's openness to the world and, more broadly, may be particularly partial to globalization in general.

Similarly, economics, marketization, and culture are covered most commonly in positive tones that have only grown more glowing over time. Positive articles on these topics began in the 1970s with China–US Ping–Pong diplomacy, and eventually comprise 1/4 to 1/2 of articles on these three topics, the remainder of which are mostly neutral articles. This agrees with the intuition that most Americans like Chinese culture. The New York Times has been deeply enamored with Chinese cultural products ranging from Chinese art to Chinese food since the very beginning of our sample. Following China's economic reforms, the number of positive articles and the proportion of positive articles relative to negative articles increases for both economics and marketization.

In contrast, welfare and well-being are covered in an almost exclusively negative light. About 1/4 of the articles on this topic are negative, and almost no articles on this topic are positive. Topics regarding politics are covered very negatively. Negative articles on ideology, government and administration, and democracy outnumber positive articles on these topics for all of the years in our sample. Though small fluctuations that coincided with ebbs in US–China relations are observed for those three topics, coverage has only grown more negative over time. Government and administration is the only negatively covered topic that does feature some positive articles. This reflects the qualitative understanding that The New York Times thinks that the Chinese state is an unpleasant but capable actor.

Despite the remarkable diversity of sentiment toward China across the eight topics, sentiment within each of the topics is startlingly consistent over time. This consistency attests to the incredible stability of American stereotypes towards China. If there is any trend to be found here, it is that the main direction of sentiment in each topic, positive or negative, has grown more prevalent since the 1970s. This is to say that reporting on China has become more polarized, which is reflective of broader trends of media polarization (Jacobs and Shapiro, 2011; Mullainathan and Shleifer, 2005).
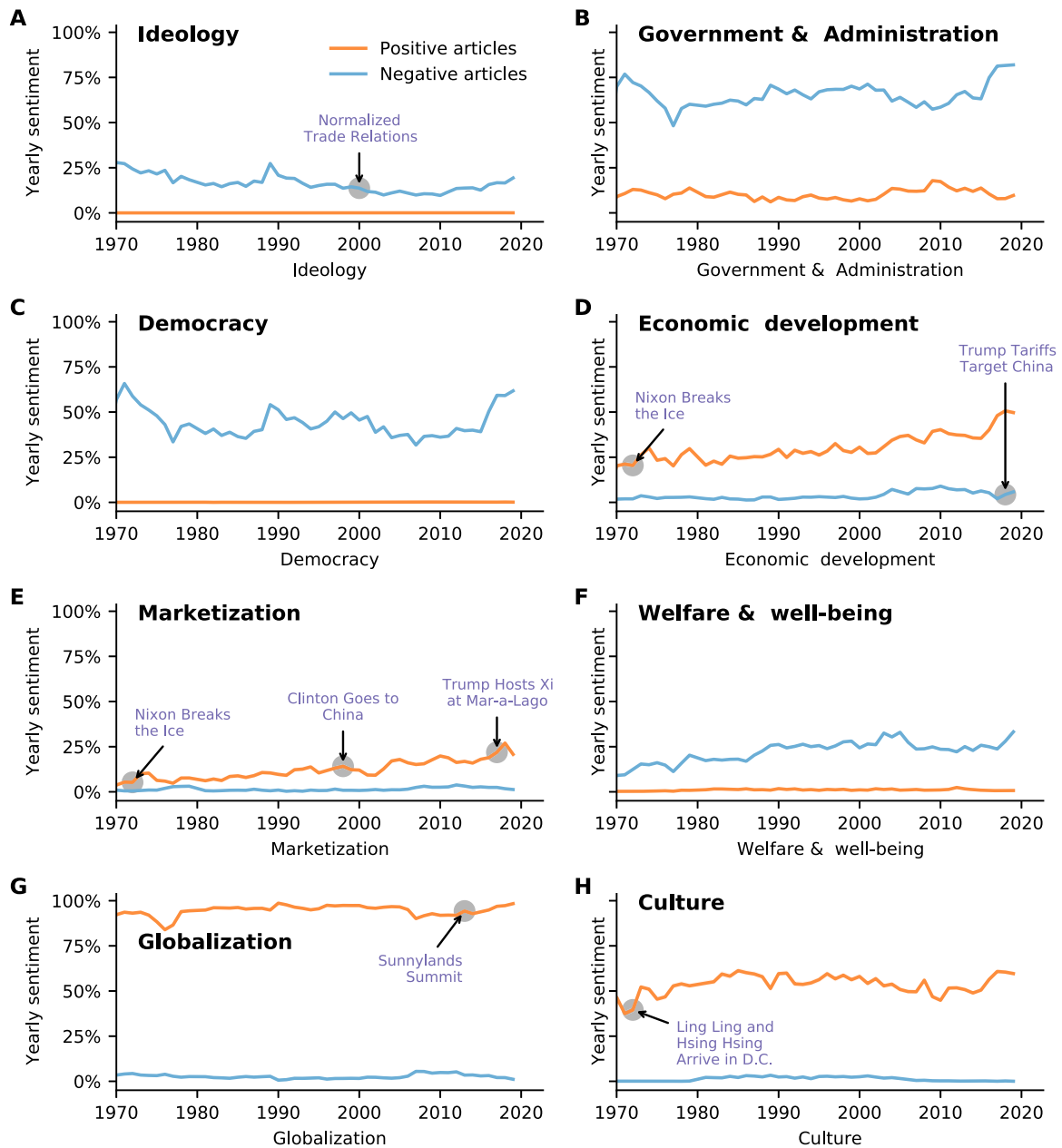
Fig. 3 Sentiments on The New York Times on China in eight topics from 1970 to 2019. The panel reports the trend of yearly media attitude toward China in (**A**) ideology, (**B**) government & administration, (**C**) democracy, (**D**) economic development, (**E**) marketization, (**F**) welfare & well-being, (**G**) globalization, and (**H**) culture. The media attitude is measured as the percentages of positive articles and negative articles, respectively. US–China relation milestones are marked as gray dots. The New York Times express diverging but consistent attitudes in the eight domains, with negative articles consistently common in ideology, government, democracy, and welfare, and positive sentiments common in economic, globalization, and culture. Standard errors are too small to be visible (below 1.55% in all topics all years).

Media sentiment affects public opinion. To reveal the connection between media sentiment and public opinion, we run a linear regression model (Eq. (1)) to fit public opinion with media sentiment from current and preceding years.

$$\mu_t = \sum_{1 \le k \le 8} \sum_{j \in [t, t-1, t-2, \cdots]} \sum_{s \in \{\text{positive,negative}\}} \beta_{kjs} F_{kjs}, \quad (1)$$

where μ_t denotes public opinion in year t with possible values ranging from -1 to 1. F_{kjs} is the fraction of positive ($s = $ positive) or negative ($s = $ negative) articles on topic k in year j. Coefficient β_{kjs} quantifies the importance of F_{kjs} in predicting μ_t.

There is inertia to public opinion. A broadly held opinion is hard to change in the short term, and it may require a while for

media sentiment to affect how the public views a given issue. For this reason, j is allowed to take $[t, t-1, t-2, \ldots]$ anywhere from zero to a couple of years ahead of t. In other words, we inspect lagged values of media sentiment as candidate predictors for public attitudes towards China.

We seek an optimal solution of media sentiment predictors to explain the largest fraction of variance (r^2) of public opinion. To reduce the risk of overfitting, we first constrain the coefficients to be non-negative after reverse-coding negative sentiment variables, which means we assume that positive articles have either no impact or positive impact and that negative articles have either zero or negative impact on public opinion. Secondly, we require that the solution be sparse and contain no more than

Table 2 Nested models fitting public opinion with media sentiment.

	1 topic	2 topics	3 topics	4 topics	5 topics	6 topics	7 topics	8 topics
$F_{culture,t-1,positive}$	2.479	3.112	2.233	2.044	1.826	1.551	1.082	1.070
$F_{democracy,t-1,negative}$		1.452	1.498	1.290	1.196	1.115	1.188	1.067
$F_{globalization,t-1,positive}$			1.945	1.951	2.047	1.926	1.963	1.818
$F_{government\&administration,t-4,negative}$				0.762	1.087	1.197	1.246	1.249
$F_{marketization,t-4,negative}$					2.869	3.816	2.791	3.186
$F_{welfare\&well-being,t-5,positive}$						5.412	8.245	8.297
$F_{economic development,t-2,negative}$							1.607	1.794
$F_{ideology,t-2,negative}$								0.324
Intercept	−1.077	−0.791	−2.136	−1.638	−1.401	−1.146	−0.846	−0.692
Explained variance	0.312	0.539	0.577	0.606	0.622	0.639	0.656	0.659

We regress public opinion on various numbers of media sentiment predictors, requiring each topic with (a) no more than one predictor, and (b) non-negative coefficients. The best two topic predictors (yearly fraction of positive articles on Chinese culture in the previous year, and yearly fraction of negative articles on Chinese democracy in the previous year) explain 53.9% of the variation in public opinion.

one non-zero coefficient in each topic:

$$\underset{\beta}{\text{maximize}} \quad r^2(\mu, \beta, F)$$

$$\text{subject to} \quad \beta_{kjs} \geq 0, \forall k, j, s$$

$$\|\beta_{k,\cdot,\cdot}\|_0 \leq 1, \forall k$$

where $r^2(\mu, \beta, F)$ is the explained variance of μ fitted with (β, F). The l_0-norm $\|\beta_{k,\cdot,\cdot}\|_0$ gives the number of non-zero coefficients of topic k predictors.

The solution varies with the number of topics included in the fitting model. As shown in Table 2, if we allow fitting with only one topic, we find that sentiment on Chinese culture has the most explanatory power, accounting for 31.2% of the variance in public opinion. We run a greedy strategy to add additional topics that yield the greatest increase in explanatory power, resulting in eight nested models (Table 2). The explanatory power of our models increases monotonically with the number of allowed topics but reaches a saturation point at which the marginal increase in variance explained per topics decreases after only two topics are introduced (see Table 2). To strike a balance between simplicity and explanatory power, we use the top two predictors, which are the positive sentiment of culture and the negative sentiment of democracy in the previous year, to build a linear predictor of public opinion that can be written as

$$\mu_t = -0.791 + 3.112F_{culture,t-1,positive} + 1.452F_{democracy,t-1,negative}, \quad (2)$$

where $F_{culture,t-1,positive}$ is the yearly fraction of positive articles on Chinese culture in year $t-1$ and $F_{democracy,t-1,negative}$ is the yearly fraction of negative articles on Chinese democracy in year $t-1$. This formula explains 53.9% of the variance of public opinion in the time series. For example, in 1993 53.9% of the articles on culture had a positive sentiment, and 46.9% of the articles on democracy had negative sentiment ($F_{culture,1993,positive} = 0.539$, $F_{democracy,1993,negative} = -0.469$). Substituting those numbers into Eq. (2) predicts public opinion in the next year (1994) to be 0.208, very close to the actual level of public opinion (0.218) (Fig. 4).

Discussion

By analyzing a corpus of 267,907 articles from The New York Times with BERT, a state-of-the-art natural language processing model, we identify major shifts in media sentiment towards China across eight topic domains over 50 years and find that media sentiment leads public opinion. Our results show that the reporting of The New York Times on culture and democracy in one year explains 53.9% of the variation in public opinion on China in the next. The conclusion that we draw from our results
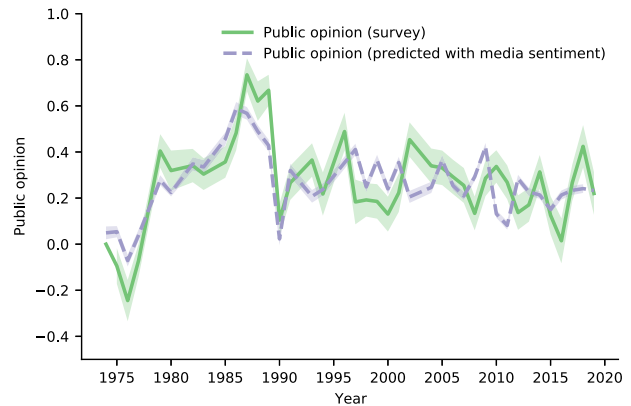
Fig. 4 Regressing public opinion of Americans toward China on The New York Times sentiments. The public opinion (solid), as a time series, is well fitted by the media sentiments on two selected topics, namely "Culture" and "Democracy", in the previous year. The dashed line shows a linear prediction based on the fractions of positive articles on "Culture" and negative articles on "Democracy" in the previous year. The public opinion is shown with a 95% confidence interval, and the fitted line is shown with one standard error.

is that media sentiment on China predicts public opinion on China. Our analysis is neither conclusive nor causal, but it is suggestive. Our results are best interpreted as a "reduced-form" description of the overall relationship between media sentiment and public opinion towards China.

While there are a number of potential factors that may complicate our conclusions, none would change the overall thrust of our results. We do not consider how the micro-level or meso-level intermediary processes through which opinion from elite media percolates to the masses below may affect our results. We also do not consider the potential ramifications of elites communing directly with the public, of major events in US–China relations causing short-term shifts in reporting, or of social media creating new channels for the diffusion of opinion. Finally, The New York Times might have a particular bias to how it covers China.

In addition to those specified above, a number of possible extensions of our work remain ripe targets for further research. Though a fully causal model of our text analysis pipeline may prove elusive (Egami et al., 2018), future work may use randomized vignettes to further our understanding of the causal effects of media exposure on attitudes towards China. Secondly, our modeling framework is deliberately simplified. The state affects news coverage before the news ever makes its way to the citizenry. It is plausible

that multiple state-level actors may bypass the media and alter public opinion directly and to different ends. For example, the actions and opinions of individual high-profile US politicians may attenuate or exaggerate the impact of state-level tension on public sentiment toward China. There are presumably a whole host of intermediary processes through which opinion from elite media affects the sentiment of the masses. Thirdly, the relationship between the sentiment of The New York Times and public opinion may be very different for hot-button social issues of first-line importance in the American culture wars. In our corpus, The New York Times has covered globalization almost entirely positively, but the 2016 election of President Donald J. Trump suggests that many Americans do not share the zeal of The Times for international commerce. We also plan to extend our measure of media sentiment to include text from other newspapers. The Guardian, a similarly elite, Anglophonic, and left-leaning paper, will make for a useful comparison case. Finally, our analysis was launched in the midst of heightened tensions between the US and China and concluded right before the outbreak of a global pandemic. Many things have changed since COVID-19. Returning to our analysis with an additional year or two of data will almost certainly provide new results of additional interest.

Future work will address some of these additional paths, but none of these elements affects the basic conclusion of this work. We find that reporting on China in one year predicts public opinion in the next. This is true for more than fifty years in our sample, and while knowledge of, for example, the opinion diffusion process on social media may add detail to this relationship, the basic flow of opinion from media to the public will not change. Regarding the putative biases of The New York Times, its ideological slant does not affect our explanation of trends in public opinion of China as long as the paper's relevant biases are relatively consistent over the time period covered by our analyses.

Data availability
All data analyzed during the current study are publicly available. The New York Times data were accessed using official online APIs (https://developer.nytimes.com/). We used their query API to search for 267,907 articles that mention China, Chinese, Beijing, Peking, or Shanghai. We downloaded the full text and date of each article. The survey data were obtained from three large public archives/centers, namely Roper Center for Public Opinion Research (ROPER), NORC at the University of Chicago, and Pew Research Center (Pew Research Center, 2019; Smith et al., 2018). See Supplementary Information for a full list of surveys. The source codes and pretrained parameters of the natural language processing model BERT are publicly released by Google Inc. on its github repository (https://github.com/google-research/bert). The fine-tuned BERT models and the inferred sentiment of The New York Times articles in our corpus are publicly available at Princeton University DataSpace. Please check the project webpage (http://www.attitudetowardchina.com/media-opinion) or the DataSpace webpage (https://doi.org/10.34770/x27d-0545) to download.

References
Aldrich J, Lu J, Kang L (2015) How do Americans view the rising China. J Contemp China 24(92):203–221

Atalay E, Phongthiengtham P, Sotelo S, Tannenbaum D (2018) New technologies and the labor market. J Monet Econ 97:48–67

Baum MA, Potter PB (2008) The relationships between mass media, public opinion, and foreign policy: toward a theoretical synthesis. Annu Rev Polit Sci 11(1):39–65

Blood DJ, Phillips PC (1995) Recession headline news, consumer sentiment, the state of the economy and presidential popularity: a time series analysis 1989–1993. Int J Public Opinion Res 7(1):2–22

Business Wire (2021) The New York Times company reports 2020 fourth-quarter and full-year results and announces dividend increase. https://www.businesswire.com/news/home/20210204005599/en/

Cao Y, Xu J (2015) The Tibet problem in the milieu of a rising China: findings from a survey on Americans' attitudes toward China. J Contemp China 24(92):240–259

Devlin J, Chang MW, Lee K, Toutanova K (2018) BERT: pre-training of deep bidirectional transformers for language understanding. Preprint at https://arxiv.org/abs/181004805

Egami N, Fong CJ, Grimmer J, Roberts ME, Stewart BM (2018) How to make causal inferences using texts. Preprint at https://arxiv.org/abs/180202163

Fiske ST, Cuddy AJ, Glick P, Xu J (2002) A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. J Personal Soc Psychol 82(6):878

Gries PH, Crowson HM (2010) Political orientation, party affiliation, and American attitudes towards China. J Chin Political Sci 15(3):219–244

Hatemi PK, McDermott R (2016) Give me attitudes. Annu Rev Political Sci 19:331–350

Hayes D, Guardino M (2011) The influence of foreign voices on US public opinion. Am J Political Sci 55(4):831–851

Iyengar S, Hahn KS, Krosnick JA, Walker J (2008) Selective exposure to campaign communication: the role of anticipated agreement and issue public membership. J Politics 70(1):186–200

Iyengar S, Kinder DR (2010) News that matters: television and American opinion, Chicago studies in American politics, updated edn. University of Chicago Press, Chicago

Jacobs L, Shapiro RY (eds) (2011) Informational interdependence: public opinion and the media in the new communications era. In: The Oxford handbook of American public opinion and the media. Oxford University Press, Oxford

Kertzer JD, Zeitzoff T (2017) A bottom-up theory of public opinion about foreign policy. Am J Political Sci 61(3):543–558

Lee E (2020) New York Times hits 7 million subscribers as digital revenue rises. https://www.nytimes.com/2020/11/05/business/media/new-york-times-q3-2020-earnings-nyt.html

Lin MH, Kwan VS, Cheung A, Fiske ST (2005) Stereotype content model explains prejudice for an envied outgroup: scale of anti-Asian American stereotypes. Personal Soc Psychol Bull 31(1):34–47

Mullainathan S, Shleifer A (2005) The market for news. Am Econ Rev 95(4):1031–1053

Peng Z (2004) Representation of China: an across time analysis of coverage in the New York Times and Los Angeles Times. Asian J Commun 14(1):53–67

Pew Research Center (2019) Global attitudes & trends. https://www.pewresearch.org/global/

Rothbaum J, Edwards A (2019) U.S. median household income was $63,179 in 2018, not significantly different from 2017. https://www.census.gov/library/stories/2019/09/us-median-household-income-not-significantly-different-from-2017.html

Schwarz D (2012) Endtimes? Crises and turmoil at the New York Times, Excelsior Editions, State University of New York Press, New York.

Smith TW, Davern M, Freese J, Morgan S (2018) General social surveys, 1972–2018. NORC, University of Chicago

Wang D, Xie Y, Huang J (2021) Latent attitude method for trend analysis with pooled survey data. Preprint at SocArXiv https://doi.org/10.31235/osf.io/atsq2

Wu HD, Stevenson RL, Chen HC, Güner ZN (2002) The conditioned impact of recession news: a time-series analysis of economic communication in the United States, 1987–1996. Int J Public Opin Res 14(1):19–36

Xie Y, Jin Y (2021) Global attitudes toward China: trends and correlates. J Contemp China

Yang YE, Liu X (2012) The 'China threat' through the lens of US print media: 1992–2006. J Contemp China 21(76):695–711

Zaller JR et al. (1992) The nature and origins of mass opinion. Cambridge University Press, Cambridge

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-021-00846-2.

Correspondence and requests for materials should be addressed to Y.X.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.