Framing Democracy:

Deciphering China's Anti-Democratic Propaganda using Word Embeddings

Patrick J. Chester*

August 23, 2024

Abstract

Autocrats have long used propaganda to maintain their grip on power, but what happens when they are forced to confront the appeal of alternative regimes? I employ word embeddings to measure whether there is a significant difference in how Chinese state media portrays democratic and non-democratic countries using three-way fixed effects regression analysis. My results show that the more democratic a country is, the more China's state media portrays its politics as chaotic and corrupt relative to two baseline news publications. This study offers novel insights into the behavior of Chinese state media, with significant implications for to our understanding of autocratic stability and the spread of democracy.

*Postdoctoral Researcher at the China Data Lab with the University of California, San Diego. Contact: patrickjchester@gmail.com Website: patrickjchester.com

Introduction

It is well documented that authoritarian regimes use their control over state media to influence public opinion. But what rhetorical strategies do they employ to shape it and at what scale? In this paper, I provide the first quantitative evidence of China's largest purveyor of international news framing the politics of foreign liberal democracies as chaotic and corrupt. To measure propaganda I use word embeddings, numeric representations of words used in similar contexts.

The existing theoretical and quantitative literature on propaganda has primarily focused on two propaganda strategies: pro-regime propaganda, which aims to persuade citizens of the regime's value and encourage continued support (Petrova 2011; Gelman et al. 2014; Gehlbach and Sonin 2014; Adena et al. 2015), and strength signaling, where the regime presents itself as resilient to discourage challenges (Edmond 2013; Huang 2015b).

However, an alternative propaganda strategy known as *negative legitimation* has been observed, where an authoritarian regime portrays alternative regime types as undesirable (Zhong 1996; Edel and Josua 2018). Survey data suggest this strategy may effectively increase support for authoritarian rulers (Huang 2015a), and experimental data indicates susceptibility to propaganda about foreign regimes (Mattingly et al. 2023). While some evidence from interviews indicate China has employed this strategy since the 1980s (Zhong 1996), quantitative evidence is lacking, and the scale of its implementation remains unknown.

Most existing measures of bias and propaganda rely on expert coders or third-party sources, limiting their flexibility and scalability for large media datasets. To address this, I use word embeddings to measure propaganda, producing a continuous metric to compare the strength of biases across text corpora in a language-neutral and coder-independent manner.

To assess China's state media's negative legitimation, I analyze over one million news articles from Xinhua News, comparing them with Taiwan's Central News Agency (CNA) and Agence France Press (AFP). Using three-way fixed-effects regression analysis, I investigate how China's state media portrays democratic countries' politics as more chaotic and corrupt than autocracies, controlling for potential confounding variables and examining the impact on political news. The results indicate a significant association between media framing and a country's democratic status, consistent with theories of propaganda (Huang 2015a).

This paper contributes significantly to the literature by providing quantitative evidence of Chinese media's negative legitimation strategy targeting liberal democracies. Moreover, it highlights the strategy's long-term usage by China's largest international news outlet, which has implications for Chinese people's perception of and demand for democracy and the stability of autocratic regimes. Additionally, the study showcases the application of word embeddings to quantify framing and propaganda, introducing new Chinese-language sentiment and topical dictionaries generated by the *conclust* algorithm.

The rest of the paper is organized as follows: I present the theoretical framework linking propaganda definitions to measurement targets and explaining why negative legitimation benefits authoritarian regimes. Next, I outline the reasons supporting the idea that Chinese media portrays the politics of democratic countries as chaotic and corrupt. Then, I detail the measurement methods and empirical strategy. The subsequent section reviews the results and assesses their robustness using different methods and model specifications. Finally, I discuss these findings in the context of the broader literature on the topic.

Theory

To what benefit is it for an authoritarian regime to change its citizens' beliefs? Also, which beliefs would a regime wish to manipulate with propaganda? A self-interested authoritarian ruler has a strong incentive to promote beliefs among their citizens that enhance regime stability. One of the chief threats to regime stability is the emergence of a popular revolution seeking to install an alternative regime type – typically liberal democracy (Acemoglu and Robinson 2006; Freeman and Quinn 2012). The regime can prevent revolution by manipulating their citizens' beliefs using propaganda and censorship.

From the citizen's perspective, there are at least three relevant beliefs that inform their decision to either remove the status quo regime or allow it to continue:[1] 1) the degree to which they will derive future benefits from the current regime; 2) the benefits they expect to receive in future periods should their country change regime types; 3) how costly they believe it would be to transition to the alternative regime type.

These beliefs inform citizens' willingness to either support the regime and help it maintain the status quo or to demand political reform. Given that the regime wishes to maintain power, it would prefer as many citizens as possible to believe that maintaining the status quo regime is preferable to transitioning to some alternative regime type. To achieve this with propaganda, the regime must either make itself appear desirable or make the alternative less so.

Much of the extant literature on propaganda has focused on pro-regime propaganda, which increases the perceived payoffs of the current regime (White, Oates, and Mcallister 2005; Adena et al. 2015; Carter and Carter 2016; Rozenas and Stukal 2019). Likewise, several studies have discussed how regimes use propaganda to make themselves appear costly to overthrow (Huang 2015b; Little 2017). However, the use of propaganda to make alternative regime types appear less desirable is poorly understood.[2] This is a significant gap in the literature because pro-regime propaganda may exhibit diminishing returns to effort as citizens learn to mistrust state media's coverage of the government (Chen and Shi 2001).

A negative legitimation strategy has a significant advantage compared to pro-regime propaganda. Namely, citizens of autocracies cannot directly experience life under foreign governments, so they may be more easily persuaded by negative coverage of liberal democracy. This advantage has been described by Gentzkow and Shapiro (2006) who found that media have more room for bias when media consumers do not receive information from the

[1]This model is a variant of the classic Downsian voter model (Downs 1957) with the addition of a cost parameter, which represents any significant cost associated with transitioning from an autocratic to a democratic regime, including the costs of repression, social stability, etc.

[2]There has been significant literature on the portrayal of political out-groups (see Adena et al. (2015)). In contrast, my research question explicitly deals with the framing of alternative regime types, which has been less closely examined in the literature.

world that can contradict their messages. This theoretical finding is supported by Mattingly et al. (2023) who find evidence suggesting that subjects may be highly susceptible to propaganda about the quality of regime types they do not live under.

China's government is well-positioned to use a negative legitimation propaganda strategy. In China, state-owned media outlets, such as Xinhua News, The People's Daily, or The Global Times, produce much of the news coverage of foreign affairs consumed in China. While commercialized Chinese media sources also exist in China, their coverage of foreign affairs broadly follows the narrative pattern of state media (Stockmann 2013). Moreover, Chinese citizens have limited access to foreign media sources, as their web addresses are blocked by China's national internet censorship apparatus, commonly called the "Great Firewall" (Stockmann 2013; King, Pan, and M. Roberts 2013). This control over their citizens' information environment allows the regime to influence their citizen's beliefs towards the efficacy of democratic institutions by portraying democratic countries in a negative light.

Hypotheses

The literature on persuasion tells us that not all framing strategies are equally effective at changing beliefs (Nelson and Oxley 1999). One rhetorical strategy that may be particularly potent for Chinese media consumers is portraying foreign democracies' politics as "chaotic". There are three reasons to think that such a strategy may be used to shape public opinion in China. First, there is a well-documented antipathy among people in mainland China against social and political instability. Second, this fear of instability appears to translate into support for the regime. Third, qualitative evidence suggests that Chinese elites have used such a strategy to maintain stability in the post-1989 era.

The limited polling and survey data of Chinese public opinion indicates that the Chinese public, by and large, has a strong preference for social stability and a fear of political disorder. In a poll of Beijing residents, Chen, Zhong, and Hillard (1997) found that when presented with a choice between living in an orderly society or a freer society that was more prone to disorder, 93% of respondents chose the former. These findings are supported by further polling performed by Chen (2004) and interviews performed by Zhong (1996). Several of the above authors speculated that this interest in political stability could be the byproduct of both regime propaganda and Chinese historical experience with multiple periods of social and political instability in the 20th century.[3]

Another reason why portraying democracy as chaotic may be a successful propaganda strategy for the Chinese government is that concerns about political stability are associated with support for the Chinese government. In their respective surveys of Chinese citizens, Chen, Zhong, and Hillard (1997) and Chen (2004) found that the greater the degree to which a respondent expressed support for political stability, the greater their support for the regime. Furthermore, consuming information about foreign political instability increases support for China's government. Huang (2015a) found that Chinese respondents who are more knowledgeable about political instability in foreign countries were more optimistic about China's future prospects, and to express trust in the Chinese government and political system. This evidence indicates that portraying foreign democracies as unstable may deliver political dividends for China's regime.

Moreover, evidence from interviews with journalists and government officials suggest that this rhetorical strategy has been used in China. In interviews with Chinese officials and residents, Zhong (1996) found that CCP media officials made an effort to portraying democracy and liberalism as a cause of political disorder while citing instability in the post-Soviet states as examples.

Given this evidence, I expect the following pattern in media produced by Chinese state-run media:

H1: *China's state media will portray democratic countries as being more chaotic than those of non-democracies relative to other media outlets without the same political objectives*

This media strategy would induce a belief among the Chinese public that democrati-

[3]Examples of social and political instability that many living Chinese people have survived include the Cultural Revolution and the 1989 Tiananmen Square Massacre.

zation would lead to political chaos, thereby decreasing their willingness to overthrow the regime.

However, other rhetorical approaches could be used to dissuade citizens in authoritarian regimes from holding positive views of democracy. For instance, autocracies could frame the politics of democratic countries as being wasteful and corrupt. There are many examples of Chinese media actors employing this rhetoric against democratic countries like the United States. One case of this occurred as a response to the Democracy Summit called by President Biden. The Global Times printed an extended editorial calling the United States a "corruption hub" in which the "US election has become a "money-burning game" of "one dollar, one vote"" (Global Times 2021). This sentiment was echoed by an editorial in Xinhua News, which speculated whether the United States was consumed by a "corruption illness" (Xinhua News 2021). These cases lead me to make the following prediction:

H2: *China's state media will describe democratic countries as being more corrupt than those of non-democracies relative to other media outlets*

Finally, if China's government wishes to instill the belief that liberal democracy, as a system of government, is less desirable than their own, they should target their negative rhetoric towards political events and actors within those countries. This ensures that Chinese consumers of this news are familiar with examples of negative political events that occurred in democratic countries; these examples could then be contrasted with positive political propaganda targeting their system of government. Accordingly, I expect the following patterns to hold for both the chaos and corruption propaganda strategies:

H3: *China's state media frame the politics of democratic countries as being chaotic relative to other outlets*

H4: *They will likewise frame the politics of democratic countries as being relatively corrupt.*

In sum, I expect that a passive news consumer in China would perceive democracies as being more chaotic and corrupt than autocratic regimes writ large. I argue that this differential is attributable to an effort by the Chinese government to persuade citizens that the status quo is preferable to any political alternatives. By understanding the rhetoric state media target to their citizens, we may better understand the factors that have shaped Chinese public opinion and their demand (or lack thereof) for political reform.

Methods

Measuring propaganda, information targeted at a consumer with the intent of changing their beliefs, has been a significant challenge for communications and social science scholars. Three primary methodologies have been used to measure propaganda in media: nominal measures in which all content produced by a publication is assumed to be propaganda (Adena et al. 2015); measures that utilize references to external ideologically charged content, such as think tanks (Groseclose and Milyo 2005; Chiang and Knight 2011); human-coder-produced content analysis in which humans assign propaganda labels to text data (Matthes and Kohring 2008; Rozenas and Stukal 2019).

While each of these methodologies have strengths, they also have limitations. Nominal classifications of media sources as being propaganda can depend on difficult to validate assumptions. For example, this approach depends on all content produced by that media source should be considered propaganda. This assumption can fail when media outlets have multiple competing objectives, such as maximizing advertising revenue while also promoting the status quo regime. Additionally, while content analysis produced by expert or crowd evaluators are considered a gold standard by researchers, they do have limitations. Namely, content analysis is prohibitively costly in both time and financial resources to implement on large-scale corpora. Finally, approaches that utilize external sources tend to be only applicable to cases where ideologically charged external references are included and, therefore, are not applicable beyond a small set of use cases.

In this study, I use word embeddings to measure propaganda. Word embeddings are a class of unsupervised machine learning models that take as inputs ordered sequences of words as they appear in natural text and output numeric vectors representing contexts in which words appear. Higher similarities between these vectors indicate that words frequently co-occur in a given text.

The approach to measuring propaganda I describe in this section draws upon lines of literature on word embeddings in the political, data, and cognitive science literature. Recent articles in political science have applied word embeddings towards measuring how politically relevant concepts are associated in text (Rodman 2020; Yang and Roberts 2021). Additionally, researchers have used word embeddings to identify racial and gender biases in large text corpora (Garg et al. 2018; Zhang et al. 2020). This study builds on the methods used by these scholars, using word embeddings to compare pairs of concepts across subcorpora to identify between-corpus biases in media content.

Measuring Propaganda

Ideally, a measure of propaganda should match existing conceptualizations in the communications and political science literature. In this section, I examine and synthesize literature on propaganda and apply it to construct a shared conceptualization of propaganda that is theoretically grounded and measurable.

One common thread in the propaganda literature is that propaganda is information presented with the intention of manipulating the beliefs of its consumer (Kenez 1985; Walton 1997; Jowett and O'Donnell 2018). This definition is in line with the political economy literature on persuasion. These models have modeled propaganda as a messaging strategy in which state media attempts to persuade the populace to support the regime (Gehlbach and Sonin 2014).

One of the greatest measurement challenges with propaganda produced by media is that it involves information asymmetry. The propagandist is aware of some set of facts about the world. Yet, it has an incentive to selectively present information in a way that leads propaganda consumers to behave in a way that is beneficial to the propagandist. In an authoritarian context, this can involve supporting the government, abstaining from protests against the government, or complying with mobilization campaigns (Gehlbach and Sonin 2014).

As it pertains to propaganda in the media, this implies the potential existence of two messages from a propagandist: $S^{'}_{oa}$ and S^{*}_{oa}, where the former is the manipulated association between a particular target object, o, and attribute, a, that is intended to bring about the desired belief, and the latter is the counter-factual association that would have been generated absent a desire to persuade. Were we to imagine that the associations between attributes and objects were represented numerically, the difference in the association between $S^{'}_{oa}$ and S^{*}_{oa} would represent the change in the message that is brought about by the propagandist's willful action. I call this difference the *propaganda effect*:

$$\texttt{Propaganda Effect}_{oa} = S^{'}_{oa} - S^{*}_{oa} \tag{1}$$

In other words, for a given message, the propaganda effect is the deviation in the association between a particular target object and attribute from the one that would exist absent any political motivation to persuade the message consumer.[4] The propaganized relationship between attributes and objects is easily observable given media produced by a propagandist; however, the latter must be inferred from external sources of information.

Associations between objects and attributes in text can be produced through a combination of framing, i.e. associating concepts in the text itself, and censorship. An example of the former was found by Rozenas and Stukal (2019): Russian state media was found to associate economic failures with external factors and attribute economic successes to the

[4]There is a conceptual difference between what I describe as a propaganda effect and the common use of the word "propaganda." The former is commonly used to describe messages intended to persuade. For instance, one might refer to a news article produced by Xinhua or China's Global Times as an example of propaganda. In contrast, the propaganda effect is the difference in framing observed between a propagandized message and one produced without the intent to persuade. This term is not used to describe the message itself, but instead the persuasive component of the message.

regime's policies. Censorship can also reduce associations in text that may be perceived as harmful to the government, such as criticism targeted towards the government on social media King, Pan, and M. Roberts (2013).

What external sources of information could be used to infer the counterfactual message? Models of media behavior suggest that private media tend to produce content that maximizes media consumption. This will lead to news content that is largely consistent with the ground truth as they perceive it, as factual information is valuable to consumers (Gehlbach and Sonin 2014). Of course, models of media behavior are imperfect; for instance, they do not take into account the well-documented phenomenon that market segmentation can lead to biased coverage, even in the absence of government intervention. I address this concern by selecting comparison media outlets that satisfy three criteria: 1) a reputation for unbiasedness; 2) either private ownership or public funding that is not conditional on content; 3) it produces content in an uncensored, i.e. global, media environment.

Figure 1: Concept Similarity Generation Procedure

$$\begin{tikzcd}
\mathtt{a}_1^p \arrow[r, "(1)"] & \mathtt{a}_{1,g_1}^p \arrow[r, "(2)"] & \mathtt{W}^p \arrow[r, "(3)"] & \mathtt{S}_{oa}^p \\
\mathtt{a}_2^p & \mathtt{a}_{2,g_1}^p & \mathtt{D}_{oa} \arrow[ur] & \\
\vdots & \vdots & & \\
\mathtt{a}_n^p & \mathtt{a}_{n,g_k}^p & &
\end{tikzcd}$$

To measure the propaganda effect and test hypotheses 1-4, I first need measures for $S_{oa}^{'}$ and S_{oa}^{*}, or the similarity between relevant objects and attributes for propagandized and unpropagandized news corpora. However, dyadic comparisons between objects and attributes are insufficient to test my hypotheses, as each hypothesis implies the presence of a conditioning or grouping variable; for hypotheses 1 and 2, I am interested in the association between countries and attributes over time (`S(Country, Attribute | Time)`), while for hypotheses 3 and 4, I am concerned with the association of an attribute with the politics of a given country (`S(Country, Attribute | Country)`). Because each of these strategies imply not just an object and attribute, but also a conditioning variable (S(O,A|G)), a simple comparison of similarities between an object, attribute pair is insufficient to test my hypotheses. Instead, I incorporate conditins into my analysis through restructuring the data. Figure 1 describes this measurement strategy.

In Part 1 (see Figure 1), articles are assigned grouping labels (g_k): the quarter (i.e., periods of three months) for testing hypotheses 1 and 2, and country for testing hypotheses 3 and 4. For instance, to test hypothesis 1, I group examine the similarity between each pairing of a country and chaos within each publication-quarter subcorpus (ex. `S(Argentina, Chaos | 2005-Q1)`). In contrast, to test hypotheses 3 and 4, the target object is politics and the grouping variable is the country; for example, `S(Politics, Chaos | Afghanistan)`.

To assign country labels to articles, I used a dictionary of Chinese-language country names and a majority rule: an article is labeled according to which country was mentioned more times than any other. For instance, if the United States was mentioned three times in a given article while China was mentioned twice, the article would be assigned a "USA" label.[5][6][7]

For Part 2 of my measurement strategy (see Figure 1 Part 2), skip-gram word embedding models are fit upon each publication-group subcorpus of articles. The output is a set of fitted word embedding models, $\mathtt{W}^p$, in which each element, $w^p_{gi} \in \mathtt{W}^p$, is an embedding model that was

[5] I use a dictionary of 259 terms that represent 201 unique country labels with both Taiwanese and mainland name variants compiled from various sources to assign country labels to articles.

[6] In Online Appendix C, I analyze the accuracy of this classification methodology by comparing country labels assigned by the dictionary plurality rule and two other assignment rules with country labels assigned to 2000 articles by a research assistant. I find that compared to alternative dictionary-based rules, the plurality rule had comparable accuracy with no loss in the number of articles classified. Accordingly, it is the rule used to assign country labels in the analysis for hypotheses 3 and 4.

[7] In this analysis, I am mindful of a core tradeoff implicit in the word embedding analysis method: the need for embeddings reflecting localized meanings of words and fitting good quality embeddings. The literature indicates that embedding models fit on fewer than 200 to 500 articles tend to reflect poorly the underlying semantic relationships in text (Roberts 2016; Rodman 2020; Zhou, Ethayarajh, and Jurafsky 2021). To prevent this structural relationship from biasing my findings, I directly control for the number of tokens associated with objects and attributes and exclude groups from my analysis for which any publication had produced fewer than 400 articles.

fit on a subcorpus of articles, $a_{1,gi}, a_{2,gi}, \ldots a_{n,gi}$, corresponding to the members of each group for a given publication. Each model is estimated with context windows of size 10 and 300 word embedding dimensions on the top ten thousand most frequent features of the corpus. These parameter choices were made to maximize model accuracy while minimizing training time based on the analysis performed by Spirling and Rodriguez (2021). The skip-gram word embedding model version used in this study is from the *word2vec* package implemented in Python 3.[8]

Next, we use the embedding models and a set of keywords to compute how similar concepts of interest are to one another within each subcorpus. More precisely, the word embedding models and a set of target and attribute word pairs, $\mathtt{D}_{oa}$, are used to generate a cosine similarity matrix in which every element is the cosine similarity score between a word pair from the attribute and target object within a publication-group subcorpus.[9]

In the final stage (see Figure 1 Part 3), I aggregate the elements of each similarity matrix to identify a single measure of the degree of similarity between a particular target object - attribute pair within each publication-group, or $s^p_{oa,gi} \in \mathtt{S}^p_{oa}$. I do so using the average across all term pairs grouped by concept pair and weighted by term frequency. Compared to the simple mean, the weighted average has the advantage of implicitly controlling for term frequency, i.e. giving weight to frequently occurring term pairs and less weight to those that rarely occur.[10][11]

[8] The Chinese characters are all converted to simplified to ensure that the feature sets from the Taiwanese corpus are comparable with those of the Xinhua corpus.

[9] To address concerns raised by Rodman (2020) about the sensitivity of embedding models to the inclusion of individual documents, the ordering of articles in the training data, and to the random seed used by the word embedding algorithm, I employ the bootstrapping approach used by Rodman (2020) for all cosine similarity estimates.

[10] The unweighted average is used as a robustness check to ensure that the results are not contingent on this particular aggregation method.

[11] A detailed definition of the weighted average measure of concept similarity and a discussion of its advantages over the simple mean is available in Online Appendix E.

Dictionary Generation

A requirement for estimating a measure of propaganda is a set of dictionaries representing target objects and attributes of interest:

- *Objects:* country names (H1,2); politics (H3,4); sports (placebo for H3,4)
- *Attributes:* chaos (H1,3), corruption (H2,4)

No freely available Chinese-language dictionaries of these concepts exist, so I use seed words and the *conclust* R package to generate them.[12] The rationale of having robust dictionaries to represent concepts is that they will represent the concept of interest with less error; i.e., while individual words may deviate from the intended meaning in a given subcorpus, a dictionary of conceptually related words should see less random measurement error. The sports target object was chosen to be a placebo because there is no clear strategic rationale for China's state media or the benchmark media outlets to portray other countries' sports as being more corrupt or chaotic.

Data

Two types of data are needed to measure the propaganda effect : messages produced by a propagandized media outlet and by at least one baseline media agency, the former represented by China's Xinhua News and the latter by Taiwan's Central News Agency and Agence France Press. The news content for these publications was gathered from the fifth edition of the Chinese Gigaword from Parker et al. (2011). The archive included 772,000, 690,000, and 135,000 news articles covering international affairs from CNA, Xinhua, and AFP. Both CNA and Xinhua have news articles covering 1992 to 2010, while AFP only

[12] A more detailed description of the algorithm and the keywords used to generate the dictionaries as well as an analysis of the dictionaries themselves is available in Online Appendix A. The dictionaries are available in Online Appendix B. A more detailed description of the algorithm can be found in Chester (2024).

includes coverage from 2001 to 2010. Given those differences, I focus my tests of hypotheses 1 and 2 on the period of 2001 to 2010, where all three publications have overlapping coverage.[13]

Xinhua News is China's largest state-owned media outlet, with over 170 foreign bureaus and 1900 journalists. It is responsible for producing or reprinting much of the foreign news consumed in China, generating around 700 foreign news items daily (Battistella and Reporters Without Borders 2005). The magnitude of the content it generates, combined with limits on competition from foreign sources imposed by China's great firewall, ensures that Chinese citizens, directly and indirectly, consume Xinhua's publications. These facts make Xinhua an attractive media agency to examine as a tool of negative legitimation, as no other source in China has the same capacity to influence Chinese public perceptions of foreign countries.

Like Xinhua, Taiwan's CNA generates much of the Chinese-language foreign news coverage consumed by its domestic audience. While not directly managed by the Taiwanese government,[14] CNA also receives large subsidies and is legally considered Taiwan's national news agency. In addition to producing its own content, CNA partners with many other leading international news sources, such as Reuters, AP, and Agence France-Presse.[15] Finally, CNA's content has been rated as being highly factual with a slight center-left bias by the non-profit organization Media Bias Fact Check.[16]

Founded in 1835 in Paris, Agence France-Presse is one of the oldest news organizations in the world. Like Xinhua and CNA, Agence France-Presse produces Chinese-language content and is largely focused on international coverage with its content. Unlike the two other news agencies, AFP is not state-owned, though it receives income indirectly from the French government via government subscriptions to its news services (Assemblee Nationale

[13]As a robustness check, I examine whether Xinhua and CNA differ in their framing of countries over the full period of 1992-2010.

[14]In 1996, Taiwan's government passed a law making CNA a non-profit corporation, placing it outside the Taiwanese government's direct control.

[15]https://focustaiwan.tw/aboutus

[16]The same NGO rated Xinhua as having a mixed record of factual reporting due to a lack of linked sourcing and prevalent pro-government propaganda.

2012). Overall, the news reporting of AFP has been rated by multiple independent media rating agencies to be low in ideological bias and high in factual accuracy (AllSides 2020; Ad Fontes Media 2021; Media Bias/Fact Check 2023).

Empirical Strategy

With this measure of propaganda, I compare how Xinhua frames democracies relative to CNA and AFP. But identifying a difference in framing between these publications must be done cautiously. A simple difference in means hypothesis test may fail to control for many potentially confounding variables at the country level. For instance, when countries experience increases in inflation, coverage of them may include more mentions of economic "chaos." Should Xinhua give preferential coverage to inflation, and if inflation is associated with regime type, this could bias a naive difference in means or OLS regression that failed to account for them. Therefore, to test hypotheses 1 and 2, I employ three-way fixed effects regression with controls for potentially confounding variables:

$$Y_{pct} = \alpha + \beta_1 G_{ct} P_p + \beta_2 G_{ct} + \beta_3 P_p + \beta P_p X_{pct} + \beta X_{pct} + \gamma_c + \omega_t + \lambda_{ct} + \delta_{pt} + \epsilon_{pct} \quad (2)$$

The unit of analysis in Equation 2 is the publication (p) country (c) quarter (t). The parameter Y_{pct} represents the mean cosine similarity between concept pairs. This variable's theoretical range is $[-100, 100]$, though we do not observe any negative cosine similarity scores in practice.[17] The parameter P_p is a binary indicator variable for Xinhua News; G_{ct} is the average Polity IV score of country c for a given quarter in the 2001 and 2010 period (Marshall, Jaggers, and Gurr 2017).

The term X_{kpct} refers to a battery of control variables (see Online Appendix F: Table 16

[17] Negative cosine similarity values would imply that concepts are negatively correlated, i.e., that increased frequency of words associated with a target object results in decreased frequency of an attribute, or vice versa. While negative associations are theoretically possible, we do not observe this for any concept pair, as the smallest observed similarity score between a pair of concepts for a given country is 9%.

for summary statistics) including model-level controls, such as the logged number of articles and counts of country and chaos terms, as well as country-level variables, such population and GDP per capita obtained from the International Monetary Fund's World Economic Outlook Indicators (International Monetary Fund 2019). I also interact the controls with P_p to control for heterogeneous effects within publications. Additional controls include inflation, government revenue, government expenditures, debt, GDP growth, country-level data on imports from China and Taiwan from the Correlates of War project (Barbieri and Keshk 2016), an indicator for diplomatic relations with Taiwan (Rich 2009), as well as a count of the number of conflict events that occurred in each country as measured by ACLED (Raleigh et al. 2010), the Corruption Perceptions Index for corruption (Apaza 2009), and the presence of a bilateral alliance (Gibler 2009).[18] Country-level and time-level fixed effects are represented by γ_c and ω_t. I also include the interactions between country fixed effects and time fixed effects, λ_{ct}, as well as the interaction between publication and time, δ_{pt}.[19]

The parameter of interest in Equation 2 is β_1, as it represents the percent change in cosine similarity between an object – a country, in this case – and an attribute – either chaos or corruption – one observes for every one unit change in a country's polity score for Xinhua's news coverage relative to CNA and AFP. The similarity can be interpreted as the degree to which an attribute is ascribed to the target object. Accordingly, β_1 is a measure of the propaganda effect or the difference observed between a politicized association between concepts and a non-politicized association. If China engages in a negative legitimation strategy, I expect the estimator $\hat{\beta}_1$ to be statistically significant and positive. This would indicate that, relative to the baselines, Xinhua describes countries as being more chaotic or corrupt the more democratic they are.

[18] All control variables, except corruption, are measured at the quarter level. Unfortunately, there do not appear to be any quarterly measures of corruption.

[19] Time and country fixed effects are included to account for potential ommitted variable bias associated with individual countries and time periods. The interactions between these variables control for country-specific shocks that may be correlated with both regime type the outcome. Finally, publication-time fixed effects are included to control for publication-specific behaviors that both vary over time and are potentially correlated with the outcome varaible. Overall, I find that fixed effects do not meaningfully affect the results.

The parameter β_2 represents how the association between the target object and attribute varies according to regime type within AFP and CNA's news coverage. Given the findings of Goldstone et al. (2010) that full autocracies and full democracies tend to be the most stable regime types, while partial autocracies and democracies tend to see more instability, one would expect that there would be either a modestly negative or even non-linear relationship between a country's polity score and the baselines' association of chaotic sentiment with them. Similarly, the theoretical and quantitative literature on political corruption suggests that polities with larger winning coalitions, i.e., democracies, experience less corruption than do autocracies (Bueno de Mesquita et al. 2001; Montinola and Jackman 2002; Sung 2004; Drury, Krieckhaus, and Lusztig 2006). Given these empirical regularities and assuming that the baseline publications produce news in ways that are consistent with them, I expect β_2 to have a negative coefficient for both attributes.

Thus far, we have described how we expect the framing of countries to differ across media sources. However, to test hypotheses 3 and 4, it is necessary to introduce an alternative object of interest: "politics." This poses a challenge: how do we measure how the politics of a particular country are associated with an attribute of interest. To address this issue, I use an alternative dictionary of political terms as the object and divide the data into subsets according to the country mentioned most frequently in each given article. This approach allows us to examine to what extent the politics of a given country are framed as chaotic or corrupt. Additionally, it allows us to introduce a placebo object, sports, to examine the robustness of any findings and measurement strategy. Equation 3 describes the new model designed to test hypotheses 3 and 4.

$$Y_{pc} = \alpha + \beta_1 G_c P_p + \beta_2 G_c + \beta_3 P_p + \beta P_p X_{pc} + \beta X_{pc} + \gamma_c + \epsilon_{pc} \qquad (3)$$

Relative to Equation 2, the parameters are the same with a few significant changes. First, the unit of analysis for this model is the country (c) publication (p). Second, the dependent variable is interpreted as the cosine similarity between the object – politics or the placebo sports – and the attributes of chaos and corruption. Third, all variables, including Polity IV and the control variables, are aggregated at the country level instead of across the full 1992-2010 period. Finally, publications included in this analysis are limited to Xinhua and CNA, as AFP has approximately 20% of the corpus size, and therefore many fewer countries are covered to a large enough degree to fit meaningful embedding models on their subcorpora. As before, the main parameter of interest remains β_1, which I expect to be statistically significant and positive.

Results

In this section, I first perform several tests of my hypotheses 1 and 2 while varying the presence of control variables and fixed effects. Next, I test hypotheses 3 and 4 using the same approach. Finally, I discuss analysis of the placebo, alternative model configurations, and measurement strategies designed to determine whether my findings are robust.

In Table 1, I regress my measure of similarity between country and attribute dictionaries on the interaction between regime type and state media. The variable `Polity (IV)` represents the baseline relationship between regime type and the Country-Attribute association for CNA and AFP. In contrast, `Xinhua x Polity (IV)` represents the change in cosine similarity in Xinhua's news relative to the baselines for every one unit change in the polity index. I include eight model variations with and without control variables and country-time fixed effects and across both chaos and corruption attributes.[20]

Across all eight models, I find consistent evidence that as a country's polity score increases, Xinhua associates more chaos and corruption sentiment with them relative to CNA and AFP. In Model 1, we see that for every one unit increase in Polity IV, there is a corresponding 0.26% increase in the similarity between Politics and Chaos for Xinhua relative to

[20] In Models (1) through (4), the dependent variable is the cosine similarity between a given country and the chaos attribute, while in Models (5) through (8), the dependent variable is the similarity between each country and corruption.

Table 1: Impact of Regime Type on Association between Country labels and Negative Attributes

	Chaos				Corruption			
	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)	Model (7)	Model (8)
Independent Variables								
Xinhua	−10.22***	−38.73***			−13.25***	−24.57***		
	(0.40)	(3.61)			(0.40)	(2.89)		
CNA		−3.89***				−5.42***		
		(0.61)				(0.55)		
Polity (IV)	−0.52***	−0.21***			−0.45***	−0.19***		
	(0.07)	(0.05)			(0.06)	(0.04)		
Xinhua x Polity (IV)	0.26***	0.15***	0.28***	0.13***	0.26***	0.17***	0.28***	0.16***
	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)	(0.03)	(0.05)	(0.04)
Statistics								
Observations	15 828	13 483	15 828	13 483	15 828	13 483	15 828	13 483
Conditions								
Controls	No	Yes	No	Yes	No	Yes	No	Yes
FE	No	No	Yes	Yes	No	No	Yes	Yes

* p < 0.1, ** p < 0.05, *** p < 0.01

Note:
Dependent variable is defined as the average cosine similarity between dictionaries representing countries and the respective attribute. Fixed-effects include country, year, publication-year, and country-year effects. Robust standard-errors are clustered at the country level and reported in parentheses: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$
Full table with control variables can be found in Appendix F.

CNA. This effect is significant at the 0.01 level. The inclusion of controls in Model 2 results in a decrease of the coefficient size to 16%; however, standard errors likewise decrease so the coefficient does not decrease in significance.

When we look at Models 3 and 4, we see that the inclusion of country and time fixed effects do little to change this observed relationship. In Model 4, we can see that a one point increase in Polity is associated with a 0.2% increase in the similarity between chaos and countries covered by Xinhua relative to CNA and AFP.[21] This is remarkably similar to the effect observed in Model 2, indicating that fixed effects are not confounding the observed relationship between concept similarity and Polity. Overall, this evidence is consistent with my first hypothesis, that Xinhua preferentially associates the politics of democracies with chaotic language.

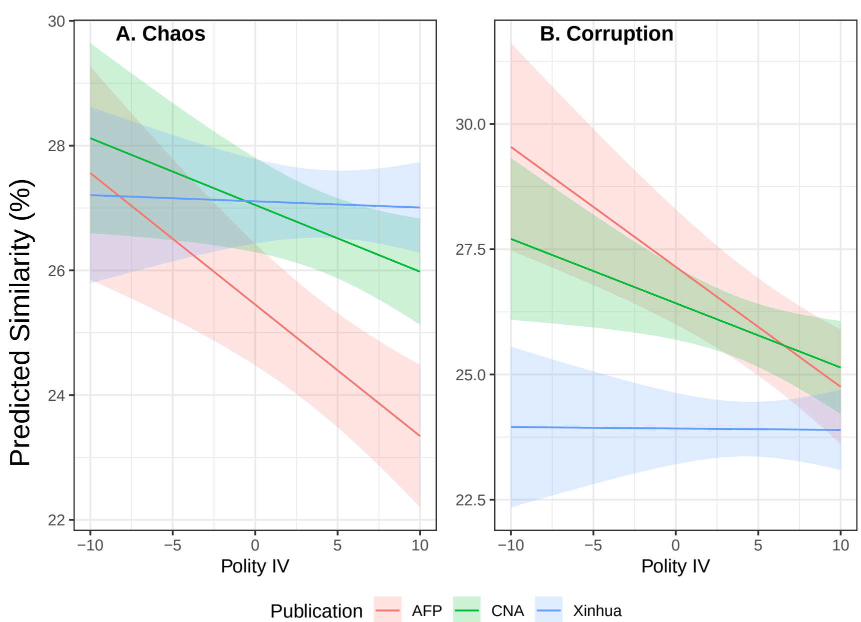
When we look at the relationship between countries and the corruption attribute, we see similar results. Across Models 5 through 8, we see a consistent pattern whereby if a country is more democratic, Xinhua associates it with higher levels of corruption relative to CNA and

[21] I use the Model 4 specification as the baseline model in subsequent regressions, as it is the most robust to omitted variable bias.

AFP. This finding is not dependent on the inclusion or exclusion of fixed effects or control variables, as p is less than 0.01 in all four cases. In sum, we see consistent evidence that supports both hypotheses 1 and 2, i.e. as Xinhua shows a willingness to portray democracies as both chaotic and corrupt relative to alternative media outlets.

To illustrate how media framing by each publication changes over various levels of Polity, I present the marginal effects of Polity on cosine similarity disaggregated by publication in Figure 2 (see Table 18 in Appendix F for full disaggregated regression results). On the y-axis we see the predicted cosine similarities between countries and attributes; on the x-axis we see each level of Polity IV represented in our data. As before, standard errors are clustered at the country level.

Figure 2: Marginal Effects of Regime Type on Country Framing



22

As with Table 1, we see differences between the slopes of each media publication, with CNA and AFP showing strongly negative slopes as Polity increases. [23] With China's state media, we see a different pattern of behavior: on an absolute scale, Xinhua covers author-

[23] This pattern is particularly pronounced for AFP, as the differential between a highly authoritarian country and a highly democratic country – with polity scores of -10 and 10, respectively – is approximately 6%. This behavior is not unexpected. As discussed before, the literature indicates that underlying corruption levels and political instability are negatively associated with levels of democracy.

itarian countries with roughly equal levels of chaotic sentiment to democratic countries. However, compared to the baseline publications, we see relatively more chaos sentiment associated with democracies. In contrast, when looking at the similarity of corruption sentiment across countries, we see differential coverage of autocracies, with Xinhua showing much lower levels of similarity.

One challenge with interpreting these findings is that it is not clear what aspects of countries are being described as "chaotic" or "corrupt." If Xinhua is engaged in a negative legitimation strategy, it will stand to reason that these attributes are targeted at the politics of democratic countries (hypotheses 3 and 4). We examine whether that is the case in the regression models presented in Table 2. For each of these models, the dependent variable is the average cosine similarity between a dictionary of terms representing "politics" and "chaos" or "corruption" attributes for a given publication-country subcorpus. To determine to what degree the results are influenced by control variables and fixed effects, they are sequentially included and excluded within each object-attribute pairing.

Table 2: Impact of Regime Type on Association between Politics Object and Negative Attributes

	Chaos				Corruption			
	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)	Model (7)	Model (8)
Independent Variables								
Xinhua	−11.55***	−39.16***	−11.55***	−16.41	−10.58***	−16.37	−10.44***	1.50
	(1.82)	(10.74)	(1.81)	(9.96)	(1.77)	(12.93)	(1.84)	(10.59)
Xinhua x Polity (IV)	0.58**	0.77***	0.58**	0.77***	0.78***	0.43*	0.76***	0.50**
	(0.24)	(0.20)	(0.24)	(0.15)	(0.23)	(0.25)	(0.24)	(0.23)
Statistics								
Observations	180	126	180	126	179	126	179	126
Effects								
Country	No	No	Yes	Yes	No	No	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note:
Dependent variable is defined as the average cosine similarity between dictionaries representing the object – politics – and the respective negative attributes. Robust standard-errors are clustered at the country level and reported in parentheses. Full table with control variables can be found in Appendix F.

Collectively, these findings indicate that Xinhua frames the politics of democracies as being more chaotic and corrupt relative to CNA. In the case of the Politics-Chaos pairing, this effect is quite significant, with a p-value of less than 0.01 and a coefficient indicating that for every one point increase in Polity, we see a 0.77 percent increase in similarity between politics and similarity terms. Across the full range of Polity, this indicates a 16% increase in similarity. Looking at the Politics-Corruption attribute-object pair, we see a smaller but still significant ($p < 0.05$) increase in similarity as Polity increases. Overall, these patterns are consistent with the hypothesis that China's state media is using a negative legitimation strategy: a news consumer who exclusively reads Xinhua News would come away with an impression that the politics of democratic countries are far more chaotic than would a reader of CNA or AFP.

Robustness

The results described above involve using a novel word embeddings-based measure of propaganda for which I had to make several potentially significant assumptions. To what extent is the measure for propaganda valid? Do these results hold when my assumptions are relaxed or when alternative measures are used? This section examines two validation checks and whether my findings are robust to alternative measurement configurations.

Validation

One potential criticism for the validity of my test of hypotheses three and four is that China's state media may cover all affairs in democratic countries in a negative light, not just their politics. To address this concern, I use a placebo object, "sports," as an alternative to "politics." Given the negative legitimation hypothesis, we have no reason to expect differential use of chaos and corruption sentiment towards sports news across publications and regime types. Accordingly, in Table 17 in Appendix F, I estimated four separate OLS models in which the dependent variables were the similarity of four attribute-object pairs: Politics-Chaos, Politics-Corruption, Sports-Chaos, Sports-Corruption. As expected, I see no statistically significant difference in how Xinhua and CNA frame sports across regime types.

Next, I validate my findings for hypotheses 1 and 2 through the use of an alternative measurement strategy: using a measure of net sentiment as a proxy for media framing.

This approach is less precise than word embeddings because it is not useful for identifying specific framing strategies. However, I expect it to be directionally consistent with the word embeddings approach. In sum, we see results consistent with expectations: when covering countries that are more democratic, Xinhua uses more negative and less positive sentiment than CNA and AFP. This analysis approach and the results are discussed in greater detail in Appendix D.

Alternative Specifications

One potential concern with the results presented in Table 1 is that this effect is driven entirely by one of the baselines, AFP or CNA. To address this concern, I disaggregated this model, running separate analyses using CNA and AFP as separate baselines for Xinhua (see Table 18 in Appendix F and Figure 2). I find that `Polity x Xinhua` remains statistically significant, whichever publication is held as a baseline. Moreover, when using CNA as a baseline, `Polity x Xinhua` remains significant no matter whether the base period of 2001 or 1992 is used (see Table 26).

Additionally, one might be concerned that these findings are a byproduct of the particular dictionary terms selected to represent the objects and attributes of interest. My main results were generated using dictionaries produced by the *conclust* algorithm, which took as inputs a small selection of seed words and a word embeddings model fitted on the full Xinhua and CNA news corpus. To determine whether my findings are dependent on this particular dictionary, I examine how they vary when I use an alternative set of dictionaries obtained by applying `conclust` to Facebook's *fastText* word embeddings model[24]. I find that the results obtained using the *fastText* dictionary are consistent with those of the baseline model (see Tables 19 and 20 in Appendix F). While we see that the *fastText* shows a significant decrease in the coefficient estimates for the Politics-Chaos concept pair, the overall effect of Polity (Xinhua) is still positive and significantly different from zero at a 95% confidence level.

[24]This model was fit on the Chinese-language text from the Common Crawl corpus and Wikipedia (Bojanowski et al. 2017)

The next set of checks involve reevaluating the skip-gram embedding models using window sizes of 5, 10, and 15 words. This entails varying the number of words considered the "context" for any word and is used to fit the word embedding model. Larger windows give weight to more distant words from the target word, while smaller windows only give weight to words that appear within a narrow distance. If we found that the association between our objects and attributes was only present for larger windows, one could argue that the association between these terms is incidental. However, in Tables 21 and 22 in Appendix F, we see significance at all levels of window size the coefficients relevant to hypotheses 1, 2, 3, and 4. This tells us that the observed associations between the objects and attributes in each instance are unlikely to be due to random word cooccurrence.

For hypotheses three and four, word embedding models were fit on country-publication subcorpora; because word embedding model quality scales with the size of the corpus it was fit upon, I excluded country-publication subcorpora from my analysis that had fewer than 400 articles, as this is in the minimum range recommended by Rodman (2020).[25] As the specific threshold I used in my analysis is somewhat arbitrary, I also examine how varying this threshold from 300 to 600 articles influences my results. The results of this analysis can be found in Table 23 in Appendix F. I find that increasing the minimum article number threshold does not meaningfully change the `Xinhua x Polity` coefficient estimates for the Politics-Chaos concept pair. However, there is a reduction in the magnitude of the Polity coefficient to insignificance for the Politics-Corruption concept pair when the minimum article threshold is set to 500; though increasing the threshold to 600 results in marginal significance at a $p < 0.1$ level. One possible explanation is that for many democratic countries covered infrequently by Xinhua, the little coverage they receive tends to involve political corruption. In any case, these results support hypothesis 3, though it is equivocal to hypothesis 4.

Finally, I performed additional checks to determine whether my results were due to

[25] As a robustness check, I vary this threshold to see to what extent it influences my findings.

the particular measurement decisions made for my dependent variable. I used the weighted average of term similarity for my dependent variable to compute the average similarity scores between concepts. This measurement strategy has advantages, such as assigning a higher weight to more frequent words. However, one could argue that the simple mean is more parsimonious. Accordingly, I check to see whether my findings still hold when using the simple mean. In Appendix F, Table 24 shows the results for hypotheses 1 and 2, while Table 25 shows the results for hypotheses 3 and 4. We see little difference compared to the baseline model for hypotheses 1, 2, and 3, with a slightly larger coefficient estimate in the latter case. However, we see a smaller coefficient estimate for `Xinhua x Polity (IV)` for the Politics-Corruption concept pair to the extent that it is not statistically significant at the 0.1 level. This result suggests that the weighted mean measure of similarity is working as intended: coefficients in the unweighted results are smaller due to greater measurement error caused by rarely occurring word pairs. Regardless, these results provide further evidence for hypotheses 1, 2, and 3, and against hypothesis 4.[26]

Overall, changing model parameters, dictionaries, and other facets of my analysis generally paint a consistent picture with those presented in my baseline model. There appears to be consistently robust evidence that the authoritarian media outlet, Xinhua News, portrays democratic politics as more chaotic than CNA and AFP, as was anticipated. Additionally, there was robust evidence that corruption sentiment is generally more strongly associated with democratic countries than authoritarian countries in Xinhua's media coverage compared to baseline outlets. However, while there was some evidence indicating that this corruption coverage was targeted toward the politics of democracies, this result appears to be sensitive to some model specifications, such as the inclusion of countries that are rarely covered and alternative measurement strategies for the dependent variable. Overall, these results present strong evidence consistent with hypotheses 1, 2, and 3 but only equivocal evidence

[26]As an additional check to ensure that these results were not driven by a single influential country, I dropped the United States from the pool of countries covered by Chinese and non-Chinese media. I find that dropping the United States does not meaningfully impact my findings. This suggests that Chinese media covers democratic countries writ large with chaotic and corrupt sentiment.

for hypothesis 4.

Conclusion

In sum, I find strong evidence consistent with the argument that China's state media tends to portray liberal democracies as chaotic and corrupt. Moreover, I have provided evidence that they target the politics of democracies with chaotic and corruption sentiment, though the association between politics and corruption appears to be dependent on a narrower range of model specifications. Overall, these findings are consistent with the predictions of the negative legitimation hypothesis: that authoritarian regimes seek to portray the politics of alternative regime types in a negative light to maintain the tacit support of their citizens. These conclusions have significant substantive and methodological implications.

First, this study has significant implications for understanding how propaganda functions in authoritarian countries. This new evidence for negative legitimation is consistent with the propaganda strategy described by Zhong (1996) and with the role that benchmarking plays in belief formation as described by Huang (2015b). However, it is notable that a similar negative portrayal of foreign countries does not appear to be employed by China's online army of social media influencers (King, Pan, and M. E. Roberts 2017). Further study is needed to determine under what conditions this strategy is employed. Moreover, it remains to be seen whether the use of a negative legitimation strategy is limited to China or is broadly practiced by other authoritarian countries.

Second, the results of this paper suggest that while Chinese state media show a consistent pattern of framing democracies as relatively more chaotic and corrupt than baselines, they appear to apply these attributes in different ways. For instance, while Xinhua tends to cover chaotic events in autocracies similar to CNA and AFP, it frames democracies as relatively more chaotic than the baselines. In contrast, authoritarian regimes are portrayed as being much less corrupt by Xinhua compared to the baselines. However, there is little difference in their portrayal of countries with high levels of democracy. This suggests that Xinhua may use negative rhetoric differently: to downplay autocratic corruption and highlight democratic chaos. This finding may have important implications for how democracy and autocracy are perceived in countries like China. It also exemplifies the advantages of word embedding-based methods as a tool to perform analysis of propaganda.

Furthermore, the procedure used by this paper to measure propaganda may have a wide variety of applications beyond the limited study of propaganda in Chinese media. Using the approach pioneered by researchers like Rodman (2020) and expanded upon in this paper, it is possible to examine to what degree political actors are using specific propaganda strategies. All that is required are dictionaries for the target object and attributes, politicized text, and a control corpus.

These findings also have implications for the resilience of authoritarian regimes writ large. In this paper, I provide evidence that Chinese state media produces media content that is consistent with a negative legitimation strategy supply of this propaganda, yet it remains to be seen whether this propaganda fulfills its purpose. Potentially, by framing alternative regime types as undesirable, autocracies have the power to shape the appeal of democracy to their citizens and thus their willingness of their citizens to challenge the regime, even when it performs poorly. Should such a link exist, it may help explain how poorly-performing autocracies, such as North Korea and Cuba, are able to prevent popular movements demanding democratic reforms from emerging. I hope that future experimental research will examine whether there exists a causal link between negative legitimation messages and skepticism towards democratic reforms among citizens of autocracies.

References

Acemoglu, Daron, and James A. Robinson. 2006. *Economic origins of dictatorship and democracy.* Cambridge University Press.

Ad Fontes Media. 2021. "Agence France-Presse Bias and Reliability." Ad Fontes Media. https://adfontesmedia.com/agence-france-presse-bias-reliability/.

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya. 2015. "Radio and the Rise of the Nazis in Prewar Germany." *The Quarterly Journal of Economics,* 1885–1939.

AllSides. 2020. "AFP Fact Check Media Bias Rating." AllSides. https://www.allsides.com/news-source/afp-fact-check-media-bias.

Apaza, Carmen R. 2009. "Measuring Governance and Corruption through the Worldwide Governance Indicators: Critiques, Responses, and Ongoing Scholarly Discussion." ISBN: 1049096509, *PS: Political Science & Politics* 42 (01): 139.

Assemblee Nationale. 2012. "Au Nom de La Commission Des Affaires Culturelles et de L' education Sur Le Projet de Loi de Finances Pour 2012." https://www.assemblee-nationale.fr/13/budget/plf2012/a3806-tvii.asp#P115_11426.

Barbieri, Katherine, and Omar M. G. Keshk. 2016. "Correlates of war project trade data set codebook, version 4.0."

Battistella, Gautier, and Reporters Without Borders. 2005. *The World's Biggest Propaganda Agency.* Technical report October.

Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. "Enriching Word Vectors with Subword Information." *Transactions of the Association for Computational Linguistics* 5:135–146.

Bueno de Mesquita, Bruce, James D. Morrow, Randolph Siverson, and Alastair Smith. 2001. “Political competition and economic growth.” *Journal of Democracy* 12:58.

Carter, Brett L, and Erin Baggott Carter. 2016. “Propaganda and Protest : Evidence from Post-Cold War Africa,” 1–30.

Chen, J. 2004. *Popular political support in urban China.* 1st. Washington D.C.: Woodrow Wilson Center Press.

Chen, Jie, Yang Zhong, and Jan William Hillard. 1997. “The Level and Sources of Popular Support for China’s Current Political Regime.” *Communist and Post-Communist Studies* 30 (1): 45–64.

Chen, Xueyi, and Tianjian Shi. 2001. “Media effects on political confidence and trust in the people’s republic of China in the post - Tiananmen period.” *East Asia* 19 (3): 84–118.

Chester, Patrick. 2024. “Embedded Lexica: Extracting Keywords from Unlabeled Corpora using Word Embeddings.”

Chiang, Chun Fang, and Brian Knight. 2011. “Media bias and influence: Evidence from newspaper endorsements.” *Review of Economic Studies* 78 (3): 795–820.

Dong, Zhendong, Qiang Dong, and Changling Hao. 2010. “HowNet and its computation of meaning.” In *Coling 2010 - 23rd International Conference on Computational Linguistics, Proceedings of the Conference.*

Downs, Anthony. 1957. *An Economic Theory of Democracy.* 1st. 1–310. New York: Harper / Row.

Drury, A. Cooper, Jonathan Krieckhaus, and Michael Lusztig. 2006. “Corruption, Democracy, and Economic Growth.” *International Political Science Review* 27 (2): 121–136.

Edel, Mirjam, and Maria Josua. 2018. "How authoritarian rulers seek to legitimize repression: framing mass killings in Egypt and Uzbekistan." *Democratization* 25 (5): 882–900.

Edmond, Chris. 2013. "Information manipulation, coordination, and regime change." *Review of Economic Studies* 80 (4): 1422–1458.

Freeman, John R., and Dennis P. Quinn. 2012. "The economic origins of democracy reconsidered." *American Political Science Review* 106 (1): 58–80.

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. "Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes." *PNAS* 115 (16): 635–644.

Gehlbach, Scott, and Konstantin Sonin. 2014. "Government control of the media." *Journal of Public Economics* 118 (October): 163–171.

Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. 2014. *Bayesian data analysis.* 661.

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

Gibler, Douglas M. 2009. *International Military Alliances, 1648-2008.* Congressional Quarterly Press.

Global Times. 2021. *America: Anti-Corruption Leader or Corruption Hub?*

Goldstone, Jack A, Robert Bates, David Lester Epstein, Ted Robert Gurr, Michael B Lustik, Monty G Marshall, Jay Ulfelder, and Mark Woodward. 2010. "A Global Model for Forecasting Political Instability." *American Journal of Political Science* 54 (1): 190–208.

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

Hamilton, William L., Kevin Clark, Jure Leskovec, and Dan Jurafsky. 2016. “Inducing domain-specific sentiment lexicons from unlabeled corpora.” *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings,* 595–605.

Huang, Haifeng. 2015a. “International Knowledge and Domestic Evaluations in a Changing Society: The Case of China.” *American Political Science Review* 109 (3): 613–634.

———. 2015b. “Propaganda as Signaling.” *Comparative Politics* 47 (4): 419–437.

International Monetary Fund. 2019. *World Economic Outlook Indicators.* Washington D.C.

Jowett, Garth S., and Victoria O’Donnell. 2018. *Propaganda and Persuasion.* Sixth. 462. Los Angeles: Sage.

Kenez, Peter. 1985. *The Birth of the Propaganda State: Soviet Methods of Mass Mobilization.* 1–324.

King, Gary, Jennifer Pan, and Margaret Roberts. 2013. “How Censorship in China Allows Government Criticism but Silences Collective Expression.” *American Political Science Review* 107 (917): 326–343.

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2017. “How the Chinese government fabricates social media posts for strategic distraction, not engaged argument.” *American Political Science Review* 111 (3): 484–501.

Lee, Sungjick, and Han Joon Kim. 2008. “News keyword extraction for topic tracking.” *Proceedings - 4th International Conference on Networked Computing and Advanced Information Management, NCM 2008* 2:554–559.

Li, Xiaotao, Shujuan You, Yawen Niu, and Wai Chen. 2021. “Learning Embeddings for Rare Words Leveraging Internet Search Engine and Spatial Location Relationships.” In *Proceedings of *SEM 2021: The Tenth Joint Conference on Lexical and Computational Semantics,* 278–287. Proceedings of SEM 2021: The Tenth Joint Conference on Lexical and Computational Semantics. Online: Association for Computational Linguistics.

Little, Andrew T. 2017. “Propaganda and credulity.” *Games and Economic Behavior* 102:224–232.

Marshall, Monty G., Keith Jaggers, and Ted Robert Gurr. 2017. “Polity IV Project. Political Regime Characteristics and Transitions, 1800-2016.” *Polity IV Project.*

Matthes, Jörg, and Matthias Kohring. 2008. “The content analysis of media frames: Toward improving reliability and validity.” *Journal of Communication* 58 (2): 258–279.

Mattingly, Daniel, Trevor Incerti, Changwook Ju, Colin Moreshead, Seiki Tanaka, and Hikaru Yamagishi. 2023. “Chinese State Media Persuades a Global Audience That the “China Model” Is Superior: Evidence From A 19-Country Experiment.”

Media Bias/Fact Check. 2023. “AFP (Agence France Presse) - Bias and Credibility.” Media Bias/Fact Check. https://mediabiasfactcheck.com/afp-agence-france-presse/.

Montinola, GR, and RW Jackman. 2002. “Sources of corruption: A cross-country study.” *British Journal of Political Science* 32:147–170.

Nelson, Thomas E, and Zoe M. Oxley. 1999. “Issue Framing Effects on Belief Importance and Opinion.” *The Journal of Politics* 61 (4): 1040–1067.

Parker, R, D Graff, K Chen, J Kong, and K Maeda. 2011. “Chinese Gigaword Fifth Edition. Linguistic Data Consortium. Philadelphia.”

Petrova, Maria. 2011. “Newspapers and parties: How advertising revenues created an independent press.” *American Political Science Review* 105 (4): 790–808.

Quinn, Kevin M., Burt L. Monroe, Michael Colaresi, Michael H. Crespin, and Dragomir R. Radev. 2010. "How to analyze political attention with minimal assumptions and costs." *American Journal of Political Science* 54 (1): 209–228.

Raleigh, Clionadh, rew Linke, Håvard Hegre, and Joakim Karlsen. 2010. "Introducing ACLED: An armed conflict location and event dataset." Publisher: Sage Publications Sage UK: London, England, *Journal of peace research* 47 (5): 651–660.

Rehurek, Radim, and Petr Sojka. 2010. "Software Framework for Topic Modelling with Large Corpora." *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks,* 45–50.

Rich, Timothy S. 2009. "Status for sale: Taiwan and the competition for diplomatic recognition." *Issues and Studies* 45 (4): 159–188.

Roberts, Kirk. 2016. "Assessing the corpus size vs. similarity trade-off for word embeddings in clinical NLP." In *Proceedings of the Clinical Natural Language Processing Workshop (ClinicalNLP),* 54–63.

Rodman, Emma. 2020. "A Timely Intervention: Tracking the Changing Meanings of Political Concepts with Word Vectors." *Political Analysis* 28:87–111.

Rothe, Sascha, Sebastian Ebert, and Hinrich Schütze. 2016. "Ultradense word embeddings by orthogonal transformation." *2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 - Proceedings of the Conference,* 767–777.

Rozenas, Arturas, and Denis Stukal. 2019. "How autocrats manipulate economic news: Evidence from Russia's state-controlled television." *Journal of Politics* 81 (3): 982–996.

Spirling, Arthur, and Pedro L Rodriguez. 2021. "Word Embeddings What works, what doesn't, and how to tell the difference for applied research." *Journal of Politics,* 1–56.

Stockmann, Daniela. 2013. *Media Commercialization and Authoritarian Rule in China.* 1st. 335. New York: Cambridge University Press.

Sung, Hung-En. 2004. "Democracy and political corruption: A cross-national comparison." *Crime, Law and Social Change* 41 (2): 179–193.

Walton, Douglas. 1997. "What Is Propaganda, and What Exactly Is Wrong with It." *Public Affairs Quarterly* 11 (4): 383–413.

White, Stephen, Sarah Oates, and Ian Mcallister. 2005. "Media effects and Russian elections, 1999-2000." *British Journal of Political Science* 35 (2): 191–208.

Xinhua News. 2021. *A Perspective on One of America's Illnesses: Legalized Corruption.*

Yang, Eddie, and Margaret E Roberts. 2021. "Censorship of Online Encyclopedias : Implications for NLP Models."

Yang, Yi, Shimei Pan, Yangqiu Song, Jie Lu, and Mercan Topkara. 2016. "Improving topic model stability for effective document exploration." In *IJCAI International Joint Conference on Artificial Intelligence,* vol. 2016-Janua, 4223–4227. 2.

Zhang, Haoran, Amy X. Lu, Mohamed Abdalla, Matthew McDermott, and Marzyeh Ghassemi. 2020. "Hurtful words: Quantifying biases in clinical contextual word embeddings." *arXiv,* 110–120.

Zhong, Yang. 1996. "Legitimacy crisis and legitimation in China." *Journal of Contemporary Asia* 26 (2): 201–220.

Zhou, Kaitlyn, Kawin Ethayarajh, and Dan Jurafsky. 2021. “Frequency-based Distortions in Contextualized Word Embeddings.”

Framing Democracy

Online Appendix

Appendices

A conclust

Before identifying the existence of an association between a particular attribute and target object in text, it is first necessary to identify a measure of both concepts. One intuitive approach would be to use a dictionary of interrelated words that collectively represent the meaning associated with the target object. Accordingly, Chinese-language dictionaries for the core concepts relevant to our hypotheses, such as politics and corruption, are needed. While there are no such dictionaries freely available, multiple methods exist for identifying keywords in the extant literature. However, some of these approaches - such as topic models - give limited control to the researcher over what "topics" or keyword clusters are produced (Yang et al. 2016). Others based on deterministic algorithms, such as those that use tf-idf weights, fail to consider the context in which words are used (Lee and Kim 2008). I propose a new algorithm, `conclust`, to address this gap.

The *conclust* algorithm bears similarities to methods employed by Rothe, Ebert, and Schütze (2016) and Hamilton et al. (2016) that are designed to extract semantically similar sentiment dictionaries from word embedding models. What differentiates *conclust* from Rothe and Hamilton's models is that instead of being designed to extract dictionary pairs that represent a latent polarity in text, it identifies the maximally self-similar word set to the seed words provided by the user. Unlike Rothe and Hamilton's models, the keywords are not required to be adjectives or have any latent polarity dimension, allowing for the extraction of non-sentiment dictionaries, such as terms associated with the word "politics". In this application, *conclust* takes sets of four seed words, and a skip-gram word embeddings model fit upon the full Gigaword Xinhua and Central News Agency news corpora.[27] It then outputs four dictionaries that vary in length from 25 to 63 tokens.

[27] The seed terms used to identify dictionaries using *conclust* and the fitted embedding model can be seen in Online Appendix A Table 3.

The `conclust` function takes a fitted word embeddings model[28] and a set of seed words as inputs and estimates a similarity matrix between each pair of words in the corpus. Next, the model identifies the word most similar to the existing seed words. In each subsequent step, the next most similar word to the current set of words is added to the group. The algorithm ends once either a maximum dictionary size has been reached or no word exists in the corpus that has an average cosine similarity score to the current group members greater than a user-chosen similarity threshold. The intuition behind this function is that at each stage, it adds new members to the dictionary that are most similar to the current set of members until a user-determined co-similarity or dictionary-size threshold is satisfied.

Algorithm 1: conclust

Input: Seed words: S; Distance matrix: M; Size threshold: n;
Similarity threshold: t
Result: Keyword set: K
$K = S$;
while $|K| \geq n$ **do**
    $\bar{m} = \forall m \in M$ $\texttt{max(sim(}K, m\texttt{))}$;
    if *mean(sim(K,* $\bar{m}$*))* $\geq t$ **then**
        $K = K \cup \bar{m}$;
    else
        break;
    end
end

Four concepts of interest were targeted using the `conclust` function: politics, sports, chaos, and corruption. The former two are considered to be target objects, as they represent concepts that can and do frequently appear in the news. The latter two are attributes that target objects may possess to some degree. Politics, chaos, and corruption were selected to evaluate hypotheses 1 and 2. In contrast, sports was selected as a placebo.[29]

The seed words that were used to identify each concept are listed below (see Table 3).

[28]The specific model used in this project is the skip-gram word embeddings model as implemented in Python's `Gensim` package (Rehurek and Sojka 2010). In a skip-gram model, a feed-forward neural network maximizes the probability of other words for each given word in the corpus.

[29]Sports was chosen as a placebo, because it is not clear how citizens would associate sports performance with the quality of a country's political system; therefore, an authoritarian regime would benefit little for portraying the sports of democracies as being excessively corrupt or chaotic.

The main criteria used to select terms was that each seed word should be an example of the key concept as it would be used in human language. For instance, articles that have political topics frequently describe actors such as presidents (总统) and prime ministers (总理). Likewise, examples of sports - such as soccer, basketball, and baseball - were used to identify other words used in similar contexts. To identify chaos and corruption, I used adjectives with similar definitions to one another.

Table 3: Seed words

	Seed word	English definition	Concept
1	民主	Democracy	Politics
2	选举	Election	Politics
3	宪法	Constitution	Politics
4	总统	President	Politics
5	总理	Prime Minister	Politics
6	混乱	Chaos	Chaos
7	乱象	Chaotic situation	Chaos
8	不安	Unsettled	Chaos
9	争论	Argue	Chaos
10	腐败	Corruption	Corruption
11	贪腐	Corruption	Corruption
12	受贿	Accept bribe	Corruption
13	行贿	To bribe	Corruption
14	足球	Soccer	Sports
15	篮球	Basketball	Sports
16	棒球	Baseball	Sports
17	游泳	Swimming	Sports

The skip-gram model used as an input for the `conclust` function was fit upon the 10 thousand most frequently occurring terms in the Xinhua News, Taiwan Central News Agency, and Agence France Press news corpora from the Fifth edition gigaword corpus. The model estimated the probability of observing each term within a 10-word window of each target term. This model generated a 300×10000 word embedding matrix with 300 word embedding dimensions and 10,000 columns representing each unique term used to fit the model.

From the word embedding matrix and seed terms, the `conclust` algorithm identified 61 political terms, 63 sports terms, 25 corruption terms, and 38 chaos terms (see Tables 4 - 12 in Online Appendix B). The minimum similarity threshold used to identify these terms was fixed at 35%, meaning the model added new members to the keyword set until no more words had at least an average of 35% similarity to the current set of members. This threshold was chosen so that the dictionaries would be conceptually homogeneous, as too low a threshold might include words that are beyond the scope of the target object.

I perform a basic qualitative examination of the dictionaries produced by `conclust` by examining their definitions and comparing them with those of the seed words and the target concept. Overall, these dictionaries appear to be consistent with the target concepts. The political dictionary included terms, like election (选举), ruling party (执政党), and the names of several political parties. The sports dictionary included volleyball (排球), gymnastics (体操), and many other examples of sports, while the output for chaos and corruption were virtually all either synonymous or directly related to the original seed words. In fact, there appear to be few if any words identified by the algorithm that were not in some way related to the target object.

I also perform principal component analysis to identify whether concepts that we would expect to cluster together, in fact do so. In particular, given that the terms are plotted according to the first and second principal components, one would expect two patterns to emerge. First, words from the same concept should roughly cluster together. Second, because corruption and chaos terms are more generally more likely to describe political events than sports events, one would expect that they would be more closely aligned with the cluster of political terms than sports terms. I perform this principal component analysis using the $300 \times 10,000$ dimension word embedding matrix derived used to identify the dictionary terms.

Figure 3: Concepts over First and Second Principal Components

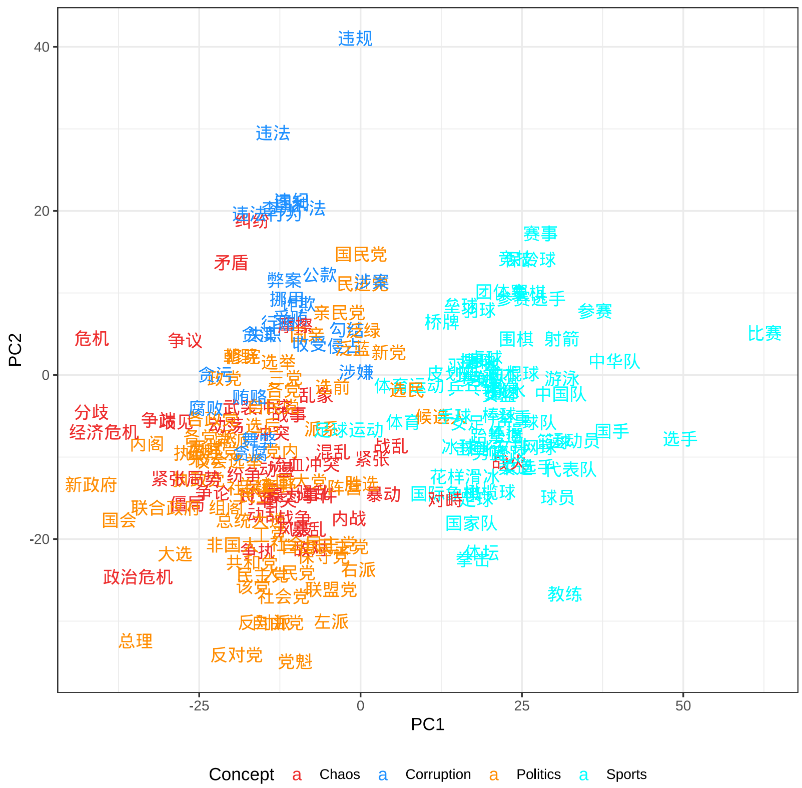


Figure 3 displays the four dictionaries plotted against their first and second principal components. The first and second principal components represent approximately 3.3% and 2.9% of variation observed in the term-word embeddings matrix. The clustering of these words conforms to expectations. First, terms in each dictionary cluster with other terms within the same dictionary. Second, the chaos and corruption terms cluster closely with the terms from the politics dictionary. These patterns also hold when the terms are evaluated over the third principal component. Generally, this suggests that descriptive words that are often to be used to describe politics tend to cluster with politics but not an unrelated concept.

In sum, I used the `conclust` function to identify four sets of keywords that are intended to represent the target objects of politics and sports and the attributes of chaos and corruption. This evidence is consistent with these keywords being reasonable approximations of the target objects and attributes.

B Chinese Dictionaries

Table 4: Keywords generated by `conclust` for Politics (1-25)

	Term	Definition	Similarity (%)
1	执政党	ruling party, the party in power	54.37
2	政党	political party	51.05
3	在野党	opposition party	49.62
4	反对党	opposition (political) party	49.24
5	该党	the party	48.78
6	两党	two party	48.34
7	在野	to be out of (political) office, to be out of power	46.85
8	民主党	Democratic Party	46.47
9	联合政府	coalition government	46.18
10	各政党	each ruling party	46.03
11	各党	each party	45.95
12	社会党	socialist party	45.66
13	三党	three parties	45.47
14	党内	within the party (esp. Chinese communist party)	45.12
15	人民党	People's party (of various countries)	44.80
16	社民党	Social Democratic party	44.75
17	大选	general election	44.20
18	选举	to elect, election	43.86
19	党派	political party, faction	43.74
20	阵营	group of people, camp, faction, sides in a dispute	43.54
21	民进党	Democratic Progressive Party (Taiwan)	43.28
22	国民党	Guomindang or Kuomintang (KMT) Nationalist Party (Taiwan)	43.00
23	自由党	Liberal Party	42.83
24	执政	to hold power, in office	42.30
25	自民党	Liberal Democratic Party (Japanese political party)	42.15

Table 5: Keywords generated by `conclust` for Politics (26-50)

	Term	Definition	Similarity (%)
26	新党	New Party (Republic of China)	41.69
27	社会民主党	Social Democratic Party	41.68
28	各党派	each party, faction	41.62
29	泛蓝	pan-blue	41.55
30	国会	Parliament (UK) Congress (US) Diet (Japan)	41.13
31	国大党	Indian Congress party	41.06
32	联盟党	Lega Nord (Italian political party)	40.81
33	组阁	to form a cabinet	40.81
34	自由民主党	Liberal Democratic Party	40.28
35	右派	(political) right, right-wing, rightist	40.14
36	保守党	conservative political parties	40.07
37	派系	sect faction	40.06
38	新政府	new government	39.75
39	工党	worker's party, labor party	39.58
40	国亲	coalition between the Taiwanese Guomindang and People's First Party	39.40
41	选后	post-election	39.37
42	反对派	opposition faction	39.07
43	总统大选	presidential election	38.90
44	选前	pre-election	38.50
45	胜选	to win an election	37.66
46	朝野	all levels of society, the imperial court and the ordinary people	37.63
47	议会选举	parliamentary or legislative election	37.61
48	选民	voter, constituency, electorate	37.48
49	修宪	to amend the constitution	37.03
50	内阁	(government) cabinet	36.82

Table 6: Keywords generated by `conclust` for Politics (51-61)

	Term	Definition	Similarity (%)
51	共和党	Republican Party	36.80
52	泛绿	pan-green	36.76
53	亲民党	People First Party (Taiwan)	36.65
54	非国大	African National Congress (South Africa)	36.51
55	候选人	candidate	36.34
56	左派	(political) left, left-wing, leftist	36.32
57	党魁	faction leader, head of political party	36.29
58	宪法	constitution (of a country)	32.97
59	民主	democracy	32.01
60	总统	president (of a country)	31.80
61	总理	premier, prime minister	28.28

Notes: Similarity (%) is the average cosine similarity of a term to other terms in the same dictionary. English definitions were obtained from the MDBG Chinese to English dictionary.

Table 7: Keywords generated by `conclust` for Sports (1-25)

	Term	Definition	Similarity (%)
1	排球	volleyball	47.68
2	射箭	archery, to shoot an arrow	46.32
3	篮球	basketball	46.02
4	跆拳道	taekwondo (Korean martial art)	45.97
5	体操	gymnastic, gymnastics	45.37
6	羽毛球	shuttlecock, badminton	45.12
7	手球	team handball	44.89
8	曲棍球	field hockey	44.34
9	桌球	table tennis, table tennis ball (Tw), billiards, pool, snooker (HK, Singapore, Malaysia)	44.32
10	垒球	softball	44.32
11	棒球	baseball	44.14
12	乒乓球	table tennis, ping-pong table, tennis ball	43.83
13	举重	to lift weights, weight-lifting (sports)	43.54
14	击剑	fencing (sport)	43.50
15	足球	soccer ball, a football, soccer, football	42.94
16	羽球	badminton	42.86
17	柔道	judo	42.84
18	保龄球	ten-pin bowling (loanword), bowling ball	42.81
19	选手	athlete, contestant	42.72
20	田径	track and field (athletics)	42.67
21	拳击	boxing	42.06
22	橄榄球	football played with oval-shaped ball (rugby, American football, Australian rules etc)	41.80
23	竞技	competition of skill (e.g. sports), athletics, tournament	41.79
24	网球	tennis, tennis ball	41.71
25	撞球	billiards, billiards ball, pool (game)	41.34

Table 8: Keywords generated by `conclust` for Sports (26-50)

	Term	Definition	Similarity (%)
26	赛艇	boat race, racing ship or boat, rowing (sport)	41.07
27	国手	(sports) member of the national team, national representative (medicine, chess etc)	40.82
28	女足	women's soccer	40.72
29	中华队	Chinese team	40.66
30	比赛	competition (sports etc), match, to compete	40.66
31	中国队	China's team	40.64
32	运动员	athlete	40.38
33	女队	women's team	40.21
34	男队	men's team	40.07
35	女排	women's volleyball abbr. for 女子排球	39.96
36	国家队	the national team	39.73
37	球队	sports team (basketball, soccer, football etc)	39.60
38	游泳	swimming, to swim	39.57
39	冰球	ice hockey, puck	39.43
40	体育	sports, physical education	39.40
41	教练	instructor, sports coach, trainer	39.29
42	代表队	delegation	39.28
43	体育运动	sports, physical culture	39.22
44	女篮	women's basketball	39.20
45	摔跤	to trip and fall, to wrestle wrestling (sports)	38.29
46	女选手	female player	37.57
47	赛事	competition (e.g. sporting)	37.49
48	皮划艇	canoe, kayak	37.49
49	桥牌	contract bridge (card game)	37.44
50	球员	sports club member: footballer, golfer etc	36.88

Table 9: Keywords generated by `conclust` for Sports (51-63)

	Term	Definition	Similarity (%)
51	围棋	the game of Go	36.84
52	国际象棋	chess	36.73
53	男篮	men's basketball, men's basketball team	36.67
54	参赛	to compete, to take part in a competition	36.66
55	武术	military skill or technique (in former times), all kinds of martial art sports (some claiming spiritual development)	36.52
56	团体赛	team competition	36.35
57	参赛选手	contestant	36.09
58	象棋	Chinese chess	35.93
59	体坛	sporting circles the world of sport	35.83
60	跳水	to dive (into water) (sports), diving, to commit suicide by jumping into water (fig.), (of stock prices etc) to fall dramatically	35.64
61	花样滑冰	figure skating	35.52
62	足球运动	soccer	35.27
63	速滑	speed skating	34.84

Notes: Similarity (%) is the average cosine similarity of a term to other terms in the same dictionary. English definitions were obtained from the MDBG Chinese to English dictionary.

Table 10: Keywords generated by `conclust` for Corruption (1-25)

	Term	Definition	Similarity (%)
1	受贿	to accept a bribe	51.24
2	贪渎	(of an official) corrupt and negligent of his duty	50.18
3	行贿	to bribe, to give bribes	49.25
4	贪污	to be corrupt, corruption, to embezzle	48.23
5	违法	illegal, to break the law	47.43
6	贿赂	to bribe, a bribe	45.93
7	不法	lawless, illegal, unlawful	45.14
8	挪用	to shift (funds) to (legitimately), to embezzle, to misappropriate	45.13
9	收受	to receive, to accept	42.86
10	图利	to seek one's benefit	42.27
11	公款	public money	41.48
12	涉案	(of a perpetrator, victim, weapon, sum of money etc) to be involved in the case	40.75
13	违纪	lack of discipline, to break a rule, to violate discipline, to breach a principle	40.14
14	舞弊	to engage in fraud	40.14
15	失职	to lose one's job, unemployment, dereliction of duty	40.10
16	违法行为	illegal behavior	39.85
17	腐败	corruption, to corrupt, to rot, rotten	39.66
18	涉嫌	to be a suspect (in a crime), to be suspected of	39.30
19	诈欺	fraud, deception	39.17
20	勾结	to collude with, to collaborate with, to gang up with	39.16
21	查办	to investigate and handle (a criminal case)	38.94
22	违规	to violate (rules), irregular, illegal, corrupt	38.67
23	弊案	scandal	37.71
24	侵占	to invade and occupy (territory)	37.44
25	贪腐	corruption	36.73

Notes: Similarity (%) is the average cosine similarity of a term to other terms in the same dictionary. English definitions were obtained from the MDBG Chinese to English dictionary.

Table 11: Keywords generated by `conclust` for Chaos (1-25)

	Term	Definition	Similarity (%)
1	衝突	conflict	55.28
2	纷争	to dispute	54.82
3	冲突	conflict, to conflict, clash of opposing forces, collision (of interests), contention	54.76
4	动乱	turmoil, upheaval, unrest	53.74
5	流血冲突	bloody conflict	53.50
6	政治危机	political crisis	49.54
7	武装冲突	armed conflict	49.24
8	争执	to dispute, to disagree, to argue opinionatedly, to wrangle	48.86
9	暴乱	riot, rebellion, revolt	48.49
10	危机	crisis	46.96
11	内战	civil war	45.58
12	紧张局势	tense situation	45.32
13	争端	dispute, controversy, conflict	45.25
14	战事	war, hostilities, fighting	44.52
15	动荡	unrest (social or political), turmoil, upheaval, commotion	43.57
16	骚乱	disturbance, riot, to create a disturbance	43.46
17	对立	to oppose, to set sth against, to be antagonistic	43.15
18	暴动	insurrection, rebellion	42.96
19	暴力事件	violent event	42.73
20	争议	controversy, dispute, to dispute	42.20
21	摩擦	friction, rubbing, chafing, fig. disharmony, conflict	41.05
22	对峙	to stand, opposite to, confront, confrontation	40.89
23	分歧	divergent, difference (of opinion, position) disagreement	39.88
24	争论	to argue, to debate, to contend, argument, contention, controversy, debate	39.73
25	战争	war, conflict	39.53

Table 12: Keywords generated by `conclust` for Chaos (26-38)

	Term	Definition	Similarity (%)
26	矛盾	contradiction, conflicting views, contradictory	39.00
27	战乱	chaos of war	38.89
28	纠纷	dispute	38.85
29	经济危机	economic crisis	38.19
30	风波	disturbance, crisis, disputes, restlessness	38.04
31	僵局	impasse deadlock	38.01
32	歧见	disagreement, differing interpretations	37.57
33	敌对	hostile, enemy (factions), combative	37.28
34	紧张	nervous, keyed-up, intense, tense, strained, in short supply, scarce	37.00
35	混乱	confusion, chaos, disorder	36.49
36	战火	conflagration, the fire of war	36.23
37	不安	unpeaceful, unstable, uneasy, disturbed, restless, worried	34.50
38	乱象	chaos, madness	34.04

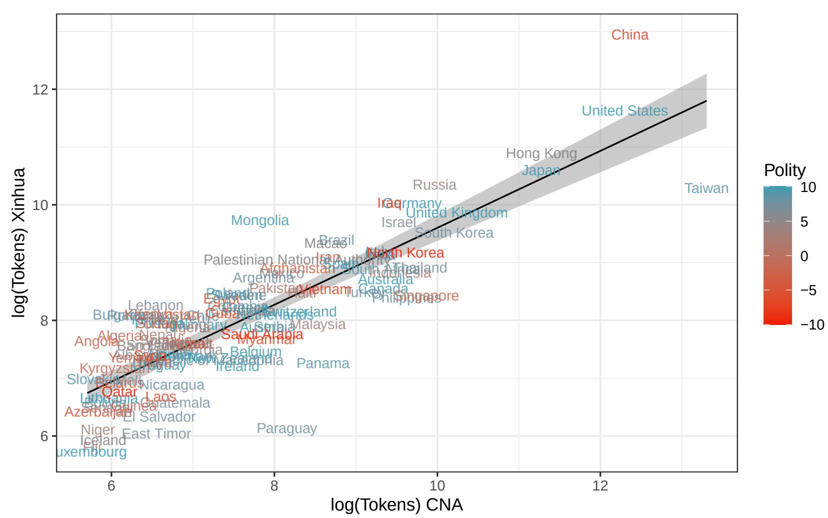
Notes: Similarity (%) is the average cosine similarity of a term to other terms in the same dictionary. English definitions were obtained from the MDBG Chinese to English dictionary.

C Country Label Assignment

Assigning high-dimensional labels to text data has been a consistent challenge in data science. Supervised machine learning models tend to struggle with this challenge, as it is very costly for human coders to label a sufficient amount of data for rarely occurring label values to be predicted with accuracy.

Assigning country labels to international news articles is a perfect example of this dilemma. There are 193 countries that the United Nations recognize, which are not discussed in the media in a symmetric way (see Figure 4). Without an impractically large training data set, a supervised machine learning model would struggle to accurately assign country labels to articles, particularly for rarely occurring labels.

Figure 4: International Text Volume Published by Publication



Given this problem, a dictionary-based approach offers some advantages. It is computationally fast and can be a powerful tool to classify documents given a set of assumptions (Quinn et al. 2010):

1. Categories are known

2. Category nesting, if any, is known

3. Relevant text features are known
4. Mapping is known
5. Coding can be automated

In this particular use case, each of these conditions is satisfied. First, there are a finite number of countries in the world; likewise, a limited number of terms are commonly used to describe them in each language. Second, countries are, for the most part, separate entities, where each word used to describe a country refers to that country alone, so there is little nesting of concern. Third, names of countries are public knowledge, accessible in practically every relevant language in publicly available databases provided by the United Nations and other international organizations.

In this section, I briefly validate several dictionary methods designed to assign country labels to news articles using a set of 2000 human-labeled news articles randomly drawn from the Chinese Gigaword 5 news corpus. I instructed the research assistant to assign labels that best describe 1) the location of the event described in the article and 2) the actors described in the news article.[30]

With these human-coded articles, I compare the performance of three different dictionary algorithms for assigning country labels to text.

1. **plurality rule:** an article is determined to be about a given country if it is mentioned more times than any other.
2. **majority rule:** an article is determined to be about a given country if at least 50% of the mentions of a country in an article are of that country; articles in which no country is mentioned a majority of the time are excluded from the analysis.

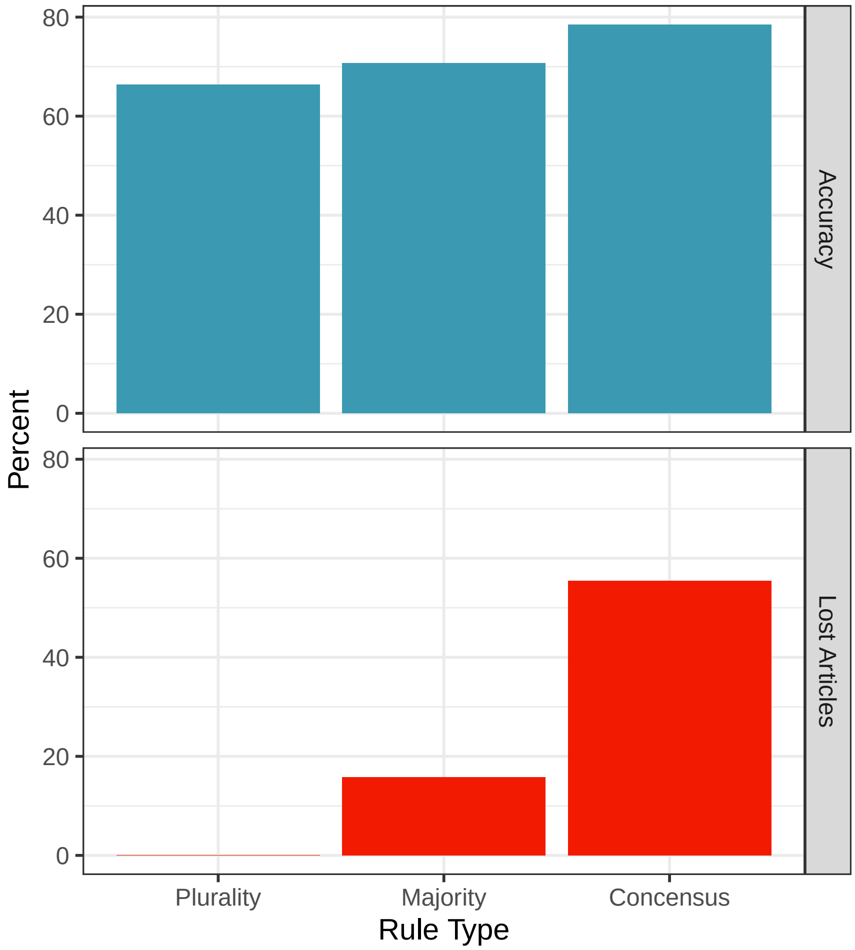
[30] I removed any mention of the publication name in each article and converted the text to simplified Chinese to decrease the potential for bias.

3. **consensus rule:** an article is determined to be about a given country if it is the only country mentioned in the article; articles in which more than one country is mentioned are excluded from the analysis.

There are two trade-offs associated with these three rules. First, stricter rules are more likely to assign country labels to articles accurately. However, the more stringent rules will also exclude articles that may have been correctly labeled and relevant to the analysis. The first trade-off is measured using classification accuracy – correctly classified articles over the total number, while the second is operationalized using the percentage of unlabeled articles using each algorithm. The following analysis aims to characterize the degree to which these trade-offs are present across each classification rule to inform which rule should be applied in my primary analysis.

In Figure 5, we can see the performance of dictionary classification across each assignment rule. The plurality rule shows a 66.35% classification accuracy rate across the 2000 labeled articles. With this rule, all documents are assigned a label. In contrast, the majority rule sees a 70.68% accuracy rate, while 15.75% of articles are unlabeled and excluded from that statistic. Finally, the consensus rule sees the highest overall accuracy, with 78.45% of articles being labeled correctly; however, 55.45% of articles received no label, making this rule the most costly of the three to implement.

Figure 5: Country Assignment Algorithm Performance



These results suggest that there indeed is a significant trade-off across the accuracy and article loss rates of these three dictionary assignment rules. The changes in accuracy across rules are non-trivial, as the most inaccurate rule (plurality) has a 12% lower classification rate than the most accurate rule (consensus); however, the differences in article loss are much more significant. The consensus rule, in particular, appears to fare a steep trade-off in unlabeled articles relative to gains in accuracy. Compared to the majority rule, it is only 7.8% more accurate but fails to label 39.7% more articles. Given these results, I use the plurality rule for my main analysis, as it offers reasonably high classification accuracy for no loss of articles.

D Sentiment Benchmark

One challenge of using word embeddings as a measurement tool is that it is quite new relative to existing methodologies. Accordingly, it is valuable to examine whether similar results can be replicated using better understood text as data methodologies. To this end, I examine the frequency of net sentiment generally, and of the attributes used in my word embedding analysis.

For my sentiment analysis I use the HowNet Chinese sentiment dictionary to quantify the average net sentiment at the subcorpus level, i.e. at the level of publication-country-quarter (Dong, Dong, and Hao 2010). I measure sentiment by taking the average net positive tokens (P_i) and negative tokens (N_i) at the level of the article, i, and take the mean of these average sentiments across all articles in a given subcorpus, N:

$$\texttt{Tone}_i = \frac{\sum_i^N P_i - N_i}{\sum_i^N Tokens_i} \tag{4}$$

In addition to measuring average net sentiment, I examine to what extent the use of positive and negative sentiment varies across publications. Accordingly, I decompose 4 into positive and negative sentiment over the total word count:

$$\texttt{Rate}_i = \frac{\sum_i^N A_i}{\sum_i^N Tokens_i} \tag{5}$$

Given hypotheses one and two, I expect that Xinhua will produce relatively negative net sentiment the more democratic a country is relative to the baseline news publications, CNA and AFP. Likewise, I expect that Xinhua's news coverage of democracies will use higher proportions of negative sentiment and lower levels of positive sentiment compared to non-democracies and the baselines.

$$Y_{pct} = \alpha + \beta_1 G_{ct} P_p + \beta_2 G_{ct} + \beta_3 P_p + \beta P_p X_{pct} + \beta X_{pct} + \gamma_c + \omega_t + \epsilon_{pct} \tag{6}$$

To test these hypotheses, I again estimate a three-way fixed-effects regression in which the units of analysis are the publication-country-quarter and the independent variable is the interaction between Xinhua (P_p) and Polity IV (G_{ct}). When using $\texttt{Tone}_{pct}$ as the dependent variable, I expect the coefficient of interest, β_1 to be negative and statistically significant. This would indicate that Xinhua covers democratic countries with a more negative tone as the country covered becomes more democratic. Conversely, when using $\texttt{Rate}_{pct}$ as the dependent variable, I expect β_1 to be negative for positive sentiment frequency and negative for negative sentiment. This would suggest that Xinhua uses relatively more negative sentiment and less positive sentiment when covering democracies. To limit risk of omitted variable bias, control variables as well as publication, country, and time fixed-effects are included in each model.

Table 13: Impact of Regime Type on Sentiment Metrics By Publication

	Tone	Rate (Positive)	Rate (Negative)
Independent Variables			
Xinhua	0.75 (1.63)	1.38 (0.61)**	−1.68 (0.41)***
CNA	3.71 (1.49)**	0.83 (0.38)**	−0.60 (0.24)**
Polity (IV)	0.03 (0.02)**	0.02 (0.01)	−0.02 (0.01)***
Xinhua x Polity (IV)	−0.06 (0.02)***	−0.03 (0.01)*	0.03 (0.01)***
Control Variables			
Country Terms (Freq)	0.09 (0.05)**	0.13 (0.04)***	0.05 (0.03)**
Attribute Term (Freq)	−0.17 (0.17)		
Imports (USD)	0.00 (0.05)	0.01 (0.04)	0.01 (0.03)
Trade Balance (USD)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
GDP (log)	0.33 (0.40)	0.05 (0.32)	−0.27 (0.27)
GDPPC (log)	−0.11 (0.37)	0.09 (0.28)	0.19 (0.27)
Inflation (%)	0.00 (0.01)	0.01 (0.01)	0.01 (0.01)
Bilateral Treaty (Count)	0.05 (0.19)	0.01 (0.14)	−0.02 (0.12)
Conflict (Freq)	0.00 (0.04)	−0.01 (0.03)	−0.02 (0.03)
Corruption Index	−0.13 (0.07)*	−0.02 (0.06)	0.11 (0.03)***
CNA x Country Terms (Freq)	−0.01 (0.04)	−0.12 (0.04)***	−0.06 (0.03)**
Xinhua x Country Terms (Freq)	−0.12 (0.06)*	−0.07 (0.05)	0.04 (0.04)
CNA x Attribute Term (Freq)	−0.28 (0.18)		
Xinhua x Attribute Term (Freq)	0.30 (0.20)		
CNA x Imports (USD)	0.09 (0.07)	0.06 (0.06)	−0.06 (0.04)
Xinhua x Imports (USD)	0.11 (0.08)	0.10 (0.07)	−0.03 (0.05)
CNA x Trade Balance (USD)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Xinhua x Trade Balance (USD)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
CNA x GDP (log)	−0.09 (0.07)	0.03 (0.07)	0.09 (0.04)**
Xinhua x GDP (log)	−0.07 (0.10)	−0.15 (0.08)*	−0.06 (0.06)
CNA x GDPPC (log)	−0.11 (0.08)	−0.06 (0.06)	0.09 (0.04)**
Xinhua x GDPPC (log)	−0.18 (0.14)	0.02 (0.10)	0.19 (0.07)***
CNA x Inflation (%)	0.00 (0.01)	−0.01 (0.01)	−0.01 (0.00)**
Xinhua x Inflation (%)	0.01 (0.01)	0.00 (0.01)	−0.01 (0.00)**
Xinhua x Bilateral Treaty (Count)	−0.02 (0.38)	0.26 (0.26)	0.31 (0.20)
CNA x Conflict (Freq)	−0.03 (0.03)	0.00 (0.03)	0.05 (0.02)**
Xinhua x Conflict (Freq)	0.00 (0.06)	−0.02 (0.04)	−0.03 (0.04)
CNA x Corruption Index	0.10 (0.04)***	0.03 (0.03)	−0.07 (0.02)***
Xinhua x Corruption Index	0.18 (0.06)***	0.06 (0.04)	−0.12 (0.04)***
Statistics			
Observations	12 013	12 013	12 013
Fixed effects			
Country	Yes	Yes	Yes
Year	Yes	Yes	Yes

Note:
In each model a separate dependent variable is used. For Tone, the outcome is the average net sentiment (P - N / Total), where P is the positive token count, N is the negative token count, and Total is the total word count associated with a given publication-country-quarter subcorpus. For Rate (Positive) and Rate (Negative) the outcome is (A / Total) where A is the total number of tokens positive or negative tokens associated with a given subcorpus. Robust standard-errors are clustered at the country level and reported in parentheses: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

Table 13 shows the results of the models described in Equation 6. In Model 1, the dependent variable is Tone, i.e. the difference in positive and negative sentiment for a given subcorpus. As expected, the coefficient associated with $Xinhua \times Polity$ is both negative and statistically significant at the 0.05 level. This suggests that Xinhua targets countries that are more democratic with more negative news coverage than they do non-democracies.

When we decompose Tone into positive and negative sentiment, we see that this effect is largely symmetric: Xinhua uses relatively more negative and less positive sentiment with democracies. The coefficients estimated in Models 1, 2, and 3 are each statistically significant at the 0.01 level.

Overall, results from this sentiment analysis are in line with my main findings: Xinhua targets democracies with negative sentiment and less positive sentiment compared to alternative news publications.

E Concept Similarity Metrics

When applying word embedding methods to data with the goal of measuring the degree to which dictionaries are similar to one another, researchers oftentimes need to compute the average similarity of word pairs across groupings of keywords or dictionaries. For instance, if a researcher wishes to know the average similarity between `politics` and `chaos`, they must aggregate across each pairing of words representing each concept. Here, let's say that politics is represented by ['government', 'president'] and chaos by ['unstable', 'chaotic']. Each of those term pairs have respective similarities as they appear in text and word frequencies, indicating how often each term appeared in the text:

Table 14: Sample Word Pair Similarity Table

Politics	Chaos	Politics (Freq)	Chaos (Freq)	Similarity (%)
government	unstable	15	8	40
government	chaotic	15	3	80
president	unstable	5	8	60
president	chaotic	5	3	75

Given that the researcher wishes to know how similar these concepts are to one another, the most obvious means of doing so is to compute the average cosine similarity across each word pair:

$$E[C] = \frac{\sum_i^N c_i}{N} \tag{7}$$

While this approach is intuitive, it has two flaws. First, not all terms are equally represented in the text. In the example above, the term 'government' occurs at three times the rate of president in the text; yet, similarities computed using the former are given equal weight to the latter using the simple mean. As a consequence, this measure may be biased by the inclusion of rarely occurring words in concept dictionaries. The second problem is that word vectors for rarely occurring words may be poorly fitted, which may result in higher rates of both random and systemic measurement error.(Li et al. 2021) To address these concerns, propose the use of a mean weighted by term frequency.

I define C_{ij} as the harmonic mean of two weighted average cosine similarity scores: $\bar{c}_i$ and $\bar{c}_j$. The first term, $\bar{c}_i$ is weighted by the frequency of terms of the attribute, and the second is weighted by the frequency of terms of the target object.

$$C_{ij} = \sqrt{\bar{c}_i \times \bar{c}_j} \tag{8}$$

where

$$\bar{c}_i = \frac{\sum_{k=1}^{n} w_{ik} c_k}{\sum_{k=1}^{n} w_{ik}} \tag{9}$$

The output of Equation 8, C_{pcij}, represents the aggregate cosine similarity between attribute i and target object j for the coverage of publication p of country c. This represents how strong an association exists for this pair of concepts in the text written about a country by a given publication.

The advantage this measure offers over the unweighted mean across all elements $m \in M$ is that the final result is weighted by the frequency of terms of both the target object and the attribute; thus, the final value will be weighted to reflect the similarity between term pairs that occur more frequently. Consequently, the weighted average similarity score would better reflect the degree to which these concepts are related from the perspective of a human reader of the text.

F Supplemental Analysis

F.1 Full Models

Table 15: Impact of Regime Type on Association between Country labels and Negative Attributes

	Chaos				Corruption			
	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)	Model (7)	Model (8)
Independent Variables								
Xinhua	−10.22 (0.40)***	−38.73 (3.61)***			−13.25 (0.40)***	−24.57 (2.89)***		
CNA		−3.89 (0.61)***				−5.42 (0.55)***		
Polity (IV)	−0.52 (0.07)***	−0.21 (0.05)***			−0.45 (0.06)***	−0.19 (0.04)***		
Xinhua x Polity (IV)	0.26 (0.05)***	0.15 (0.04)***	0.28 (0.05)***	0.13 (0.04)***	0.26 (0.05)***	0.17 (0.03)***	0.28 (0.05)***	0.16 (0.04)***
Control Variables								
Country Terms (log)		−2.00 (0.27)***		−2.74 (0.31)***		−2.87 (0.15)***		−3.22 (0.26)***
Attribute Terms (Freq)		−7.75 (0.34)***				−7.05 (0.18)***		
Imports (USD)		−0.64 (0.19)***		−0.40 (0.24)*		−0.15 (0.17)		0.10 (0.22)
Trade Balance (USD)		0.00 (0.00)		0.00 (0.00)		0.00 (0.00)		0.00 (0.00)
GDP (log)		−0.39 (0.38)				−0.41 (0.31)		
GDPPC (log)		0.54 (0.35)				0.25 (0.34)		
Inflation (%)		0.00 (0.02)				−0.05 (0.02)**		
Bilateral Treaty (Count)		−1.73 (1.28)		−3.21 (1.36)**		−2.99 (0.96)***		−3.83 (1.15)***
Conflict (Freq)		0.49 (0.23)**				0.42 (0.18)**		
Corruption Index		−0.95 (0.21)***				−0.59 (0.18)***		
Xinhua x Country Terms (log)		0.24 (0.19)		0.42 (0.41)		0.58 (0.22)***		0.94 (0.29)***
Xinhua x Attribute Terms (Freq)		4.18 (0.42)***				2.00 (0.39)***		
Xinhua x Imports (USD)		−0.61 (0.27)**		−0.69 (0.38)*		−0.23 (0.25)		−0.64 (0.35)*
Xinhua x Trade Balance (USD)		0.00 (0.00)		0.00 (0.00)		0.00 (0.00)		0.00 (0.00)
Xinhua x GDP (log)		0.90 (0.34)***		0.70 (0.47)		0.53 (0.34)		0.72 (0.46)
Xinhua x GDPPC (log)		−0.48 (0.27)*		0.26 (0.37)		−0.36 (0.27)		0.34 (0.34)
Xinhua x Inflation (%)		0.02 (0.02)		0.02 (0.03)		0.05 (0.02)**		0.02 (0.02)
Xinhua x Bilateral Treaty (Count)		1.53 (1.76)		2.84 (1.51)*		3.70 (1.32)***		4.93 (1.26)***
Xinhua x Conflict (Freq)		−0.35 (0.17)**		−0.02 (0.20)		−0.44 (0.14)***		−0.15 (0.15)
Xinhua x Corruption Index		0.50 (0.17)***		0.16 (0.21)		0.16 (0.16)		−0.15 (0.20)
Statistics								
Observations	15 828	13 483	15 828	13 483	15 828	13 483	15 828	13 483
Fixed effects								
FE	No	No	Yes	Yes	No	No	Yes	Yes

Note:
Dependent variable is defined as the average cosine similarity between dictionaries representing countries and the respective attribute. Fixed-effects include country, year, publication-year, and country-year effects. Robust standard-errors are clustered at the country level and reported in parentheses: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

F.2 Controls

Table 16: Covariate Summary Statistics and Sources

	N	Mean	SD	Min	Max	Source
Polity IV	180	3.94	6.36	−10.00	10.00	Marshall et al. (2014)
Cosine similarity (%)	180	46.56	15.43	18.89	84.73	Author (2023)
Tokens - Object (log)	180	7.95	1.38	4.57	13.69	Author (2023)
Tokens - Attribute (log)	180	6.89	1.45	3.47	11.63	Author (2023)
Imports (log)	174	4.29	2.52	0.00	10.22	Izmirlioglu (2017)
GDP (log)	176	3.59	2.07	−0.86	8.79	IMF (2019)
GDPPC (log)	174	7.73	1.70	4.28	10.59	IMF (2019)
Inflation (%)	174	150.43	567.24	−5.91	4734.91	IMF (2019)
Current Account Balance	176	−1.11	15.94	−51.61	112.39	IMF (2019)
Ally	180	0.04	0.19	0.00	1.00	Gilber (2009)
Conflict Events (log)	180	1.45	2.96	0.00	9.48	Raleigh et al. (2010)
Corruption Index	128	4.83	2.60	0.40	10.00	Transparancy International (2018)

F.3 Placebo

Table 17: Impact of Regime Type on Association between Politics Object, Sports Placebo, and Negative Attributes

	Politics		Sports	
	Chaos (1)	Corruption (2)	Chaos (3)	Corruption (4)
Independent Variables				
Xinhua	−16.41 (9.96)	1.50 (10.59)	−37.64 (10.53)***	−20.68 (10.74)*
Xinhua x Polity (IV)	0.77 (0.15)***	0.50 (0.23)**	0.50 (0.33)	−0.07 (0.31)
Control Variables				
Country Terms (Freq)	−6.42 (1.80)***	−7.97 (2.12)***	−5.89 (1.77)***	−6.20 (1.20)***
Attribute Term (Freq)	−7.69 (1.36)***	−6.96 (1.63)***	−9.00 (1.61)***	−8.57 (1.02)***
Bilateral Treaty (Count)	7.92 (4.95)	7.43 (4.68)	1.90 (5.85)	−2.66 (5.07)
Imports (USD)	−0.22 (0.34)	−0.22 (0.41)	−0.12 (0.41)	0.17 (0.50)
Xinhua x Country Terms (Freq)	0.12 (1.48)	−3.02 (1.28)**	2.73 (1.23)**	−0.42 (1.84)
Xinhua x Attribute Term (Freq)	−0.17 (1.37)	1.98 (1.14)*	0.16 (1.15)	2.49 (1.41)*
Xinhua x GDP (log)	−0.03 (0.65)	0.57 (0.95)	−2.38 (0.91)**	−1.69 (1.16)
Xinhua x GDPPC (log)	0.96 (1.19)	−0.20 (1.37)	3.12 (1.38)**	3.09 (2.19)
Xinhua x Inflation (%)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Xinhua x Trade Balance (USD)	−0.02 (0.02)	0.02 (0.04)	−0.06 (0.03)**	−0.08 (0.05)*
Xinhua x Conflict (Freq)	0.32 (0.26)	−0.05 (0.30)	0.49 (0.34)	−0.69 (0.30)**
Xinhua x Corruption Index	−0.68 (0.53)	−0.25 (0.70)	−1.03 (0.68)	−1.56 (1.09)
Statistics				
Observations	126	126	126	126
Fixed effects				
Country	Yes	Yes	Yes	Yes

Note:
Dependent variable is defined as the average cosine similarity between dictionaries representing the objects – politics and sports – and the respective attributes – chaos and corruption. Robust standard-errors are clustered at the country level and reported in parentheses: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

F.4 Alternative Baseline

Table 18: Impact of Regime Type on Association between Country labels and Negative Attributes with Varying Baseline

	Chaos			Corruption		
	AFP (1)	CNA (2)	Both (3)	AFP (4)	CNA (5)	Both (6)
Independent Variables						
Xinhua x Polity (IV)	0.20 (0.07)***	0.10 (0.05)**	0.13 (0.04)***	0.23 (0.07)***	0.14 (0.04)***	0.16 (0.04)***
Control Variables						
Country Terms (log)	−3.13 (0.43)***	−2.41 (0.34)***	−2.74 (0.31)***	−4.27 (0.35)***	−2.72 (0.29)***	−3.22 (0.26)***
Imports (USD)	−1.26 (0.41)***	−0.51 (0.34)	−0.40 (0.24)*	−0.80 (0.35)**	−0.28 (0.31)	0.10 (0.22)
Trade Balance (USD)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Bilateral Treaty (Count)	1.79 (1.08)	−0.75 (1.01)	−3.21 (1.36)**	0.57 (0.90)	1.28 (1.10)	−3.83 (1.15)***
Xinhua x Country Terms (log)	0.84 (0.46)*	−0.06 (0.46)	0.42 (0.41)	1.92 (0.48)***	0.14 (0.31)	0.94 (0.29)***
Xinhua x Imports (USD)	−0.57 (0.62)	−0.13 (0.36)	−0.69 (0.38)*	−0.36 (0.62)	−0.14 (0.34)	−0.64 (0.35)*
Xinhua x Trade Balance (USD)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Xinhua x GDP (log)	1.30 (0.74)*	−0.11 (0.46)	0.70 (0.47)	0.90 (0.75)	−0.05 (0.44)	0.72 (0.46)
Xinhua x GDPPC (log)	0.31 (0.62)	0.23 (0.35)	0.26 (0.37)	0.42 (0.53)	0.32 (0.36)	0.34 (0.34)
Xinhua x Inflation (%)	0.02 (0.04)	0.03 (0.03)	0.02 (0.03)	0.04 (0.05)	0.01 (0.02)	0.02 (0.02)
Xinhua x Bilateral Treaty (Count)	−0.20 (1.49)		2.84 (1.51)*	2.00 (1.66)		4.93 (1.26)***
Xinhua x Conflict (Freq)	−0.11 (0.34)	−0.09 (0.22)	−0.02 (0.20)	−0.15 (0.30)	−0.29 (0.20)	−0.15 (0.15)
Xinhua x Corruption Index	0.42 (0.35)	−0.08 (0.20)	0.16 (0.21)	−0.10 (0.31)	−0.23 (0.20)	−0.15 (0.20)
Statistics						
Observations	9255	9348	13 483	9255	9348	13 483
Fixed effects						
FE	Yes	Yes	Yes	Yes	Yes	Yes

Note:
Dependent variable is defined as the average cosine similarity between dictionaries representing countries and the respective attribute. Fixed-effects include country, year, publication-year, and country-year effects. Robust standard-errors are clustered at the country level and reported in parentheses: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$

F.5 Dictionary

Table 19: Impact of Regime Type on Association between Country labels and Negative Attributes Conditioning on Concept Dictionaries

	Chaos		Corruption	
	Word2Vec (1)	Fasttext (2)	Word2Vec (3)	Fasttext (4)
Independent Variables				
Xinhua x Polity (IV)	0.13 (0.04)***	0.09 (0.04)**	0.16 (0.04)***	0.16 (0.04)***
Control Variables				
Country Terms (log)	−2.74 (0.31)***	−3.35 (0.28)***	−3.22 (0.26)***	−2.81 (0.22)***
Imports (USD)	−0.40 (0.24)*	0.27 (0.22)	0.10 (0.22)	−0.57 (0.18)***
Trade Balance (USD)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Bilateral Treaty (Count)	−3.21 (1.36)**	−3.75 (1.26)***	−3.83 (1.15)***	−2.53 (1.02)**
Xinhua x Country Terms (log)	0.42 (0.41)	1.28 (0.27)***	0.94 (0.29)***	0.01 (0.29)
Xinhua x Imports (USD)	−0.69 (0.38)*	−0.74 (0.35)**	−0.64 (0.35)*	−0.50 (0.30)*
Xinhua x Trade Balance (USD)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Xinhua x GDP (log)	0.70 (0.47)	0.74 (0.41)*	0.72 (0.46)	0.73 (0.38)*
Xinhua x GDPPC (log)	0.26 (0.37)	0.94 (0.33)***	0.34 (0.34)	0.11 (0.32)
Xinhua x Inflation (%)	0.02 (0.03)	0.02 (0.03)	0.02 (0.02)	0.02 (0.02)
Xinhua x Bilateral Treaty (Count)	2.84 (1.51)*	3.73 (1.53)**	4.93 (1.26)***	2.93 (1.29)**
Xinhua x Conflict (Freq)	−0.02 (0.20)	0.08 (0.16)	−0.15 (0.15)	0.02 (0.17)
Xinhua x Corruption Index	0.16 (0.21)	−0.11 (0.20)	−0.15 (0.20)	0.06 (0.17)
Statistics				
Observations	13 483	13 396	13 483	13 494
Fixed effects				
FE	Yes	Yes	Yes	Yes

Note:
Dependent variable is defined as the average cosine similarity between dictionaries representing countries and the respective attribute. Fixed-effects include country, year, publication-year, and country-year effects. Robust standard-errors are clustered at the country level and reported in parentheses: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 20: Impact of Regime Type on Association between Politics Object and Negative Attributes by Dictionary Source

	Chaos		Corruption	
	Baseline (1)	Fasttext (2)	Baseline (3)	Fasttext (4)
Independent Variables				
Xinhua	−16.41 (9.93)	−15.09 (10.98)	1.50 (10.55)	4.12 (12.32)
Xinhua x Polity (IV)	0.77 (0.15)***	0.54 (0.18)***	0.50 (0.23)**	0.67 (0.20)***
Control Variables				
Bilateral Treaty (Count)	7.92 (4.93)	1.83 (4.48)	7.43 (4.66)	7.75 (6.21)
Imports (USD)	−0.22 (0.34)	0.35 (0.35)	−0.22 (0.41)	−0.04 (0.52)
Country Terms (Freq)	−6.42 (1.79)***	−7.48 (2.44)***	−7.97 (2.11)***	−8.78 (3.02)***
Attribute Term (Freq)	−7.69 (1.36)***	−8.52 (2.04)***	−6.96 (1.63)***	−7.19 (3.13)**
Xinhua x GDP (log)	−0.03 (0.65)	−0.16 (0.70)	0.57 (0.95)	−0.13 (1.02)
Xinhua x GDPPC (log)	0.96 (1.18)	0.17 (1.15)	−0.20 (1.36)	−0.15 (1.50)
Xinhua x Inflation (%)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Xinhua x Trade Balance (USD)	−0.02 (0.02)	0.00 (0.02)	0.02 (0.04)	−0.01 (0.02)
Xinhua x Conflict (Freq)	0.32 (0.26)	0.11 (0.27)	−0.05 (0.30)	−0.16 (0.34)
Xinhua x Corruption Index	−0.68 (0.53)	−0.50 (0.45)	−0.25 (0.70)	−0.89 (0.59)
Xinhua x Country Terms (Freq)	0.12 (1.47)	−1.01 (1.42)	−3.02 (1.27)**	−1.95 (2.00)
Xinhua x Attribute Term (Freq)	−0.17 (1.36)	1.79 (1.30)	1.98 (1.14)*	0.66 (2.10)
Statistics				
Observations	132	132	132	132
Fixed effects				
Country	Yes	Yes	Yes	Yes

Note:
Dependent variable is defined as the average cosine similarity between dictionaries representing the object – politics – and the respective attributes – chaos and corruption. Robust standard-errors are clustered at the country level and reported in parentheses: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

F.6 Window size

Table 21: Impact of Regime Type on Association between Country labels and Negative Attributes Conditioning on Window Size

	Chaos			Corruption		
	Window 5 (1)	Window 10 (2)	Window 15 (3)	Window 5 (4)	Window 10 (5)	Window 15 (6)
Independent Variables						
Xinhua x Polity (IV)	0.14 (0.05)***	0.15 (0.04)***	0.13 (0.04)***	0.12 (0.05)**	0.18 (0.04)***	0.14 (0.04)***
Control Variables						
Country Terms (log)	−4.23 (0.38)***	−3.27 (0.29)***	−2.80 (0.31)***	−4.72 (0.30)***	−4.32 (0.27)***	−2.92 (0.22)***
Imports (USD)	−0.35 (0.27)	−1.24 (0.31)***	−0.34 (0.23)	0.47 (0.28)	−0.85 (0.27)***	0.10 (0.20)
Trade Balance (USD)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Bilateral Treaty (Count)	−4.09 (1.66)**	1.40 (0.87)	−2.87 (1.28)**	−4.61 (1.56)***	0.51 (0.73)	−3.54 (1.08)***
Xinhua x Country Terms (log)	1.11 (0.43)**		0.38 (0.41)	1.66 (0.34)***		0.84 (0.29)***
Xinhua x Imports (USD)	−0.81 (0.47)*		−0.69 (0.38)*	−0.58 (0.39)		−0.58 (0.32)*
Xinhua x Trade Balance (USD)	0.00 (0.00)		0.00 (0.00)	0.00 (0.00)		0.00 (0.00)
Xinhua x GDP (log)	0.45 (0.58)		0.84 (0.46)*	0.41 (0.52)		0.63 (0.42)
Xinhua x GDPPC (log)	0.57 (0.45)		0.13 (0.36)	0.34 (0.37)		0.25 (0.31)
Xinhua x Inflation (%)	0.02 (0.03)		0.01 (0.02)	0.00 (0.02)		0.01 (0.02)
Xinhua x Bilateral Treaty (Count)	4.61 (1.89)**		3.48 (1.54)**	7.56 (1.61)***		5.95 (1.28)***
Xinhua x Conflict (Freq)	−0.04 (0.22)		−0.03 (0.20)	−0.16 (0.16)		−0.13 (0.13)
Xinhua x Corruption Index	0.02 (0.27)		0.10 (0.21)	−0.08 (0.22)		−0.17 (0.18)
Statistics						
Observations	13 384	13 483	13 398	13 478	13 483	13 487
Fixed effects						
FE	Yes	Yes	Yes	Yes	Yes	Yes

Note:
Dependent variable is defined as the average cosine similarity between dictionaries representing countries and the respective attribute. Fixed-effects include country, year, publication-year, and country-year effects. Robust standard-errors are clustered at the country level and reported in parentheses: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 22: Impact of Regime Type on Association between Politics Object and Negative Attributes

	Chaos			Corruption		
	Win 5 (1)	Win 10 (2)	Win 15 (3)	Win 5 (4)	Win 10 (5)	Win 15 (6)
Independent Variables						
Xinhua	−21.99 (10.44)**	−16.41 (9.93)	−12.67 (8.91)	−9.81 (11.49)	1.50 (10.55)	5.39 (11.28)
Xinhua x Polity (IV)	0.81 (0.14)***	0.77 (0.15)***	0.74 (0.17)***	0.45 (0.18)**	0.50 (0.23)**	0.57 (0.27)**
Control Variables						
Bilateral Treaty (Count)	5.04 (4.87)	7.92 (4.93)	9.01 (5.04)*	7.01 (3.45)**	7.43 (4.66)	8.34 (5.76)
Imports (USD)	0.07 (0.32)	−0.22 (0.34)	−0.35 (0.33)	−0.01 (0.33)	−0.22 (0.41)	−0.34 (0.46)
Country Terms (Freq)	−8.40 (1.55)***	−6.42 (1.79)***	−5.91 (1.82)***	−9.98 (2.08)***	−7.97 (2.11)***	−7.12 (2.21)***
Attribute Term (Freq)	−7.95 (1.15)***	−7.69 (1.36)***	−7.08 (1.41)***	−7.24 (1.55)***	−6.96 (1.63)***	−6.03 (1.81)***
Xinhua x GDP (log)	−0.14 (0.69)	−0.03 (0.65)	−0.03 (0.60)	−0.12 (0.85)	0.57 (0.95)	0.56 (1.08)
Xinhua x GDPPC (log)	0.37 (1.15)	0.96 (1.18)	1.07 (1.10)	0.27 (1.33)	−0.20 (1.36)	−0.24 (1.56)
Xinhua x Inflation (%)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Xinhua x Trade Balance (USD)	−0.01 (0.02)	−0.02 (0.02)	−0.02 (0.02)	0.02 (0.04)	0.02 (0.04)	0.02 (0.04)
Xinhua x Conflict (Freq)	0.47 (0.24)*	0.32 (0.26)	0.14 (0.27)	−0.03 (0.26)	−0.05 (0.30)	−0.12 (0.39)
Xinhua x Corruption Index	−0.50 (0.48)	−0.68 (0.53)	−0.72 (0.51)	−0.41 (0.60)	−0.25 (0.70)	−0.08 (0.78)
Xinhua x Country Terms (Freq)	0.15 (1.43)	0.12 (1.47)	−0.16 (1.28)	−2.13 (1.28)	−3.02 (1.27)**	−3.48 (1.53)**
Xinhua x Attribute Term (Freq)	0.74 (1.27)	−0.17 (1.36)	−0.25 (1.30)	2.39 (1.22)*	1.98 (1.14)*	2.15 (1.50)
Statistics						
Observations	132	132	132	132	132	132
Fixed effects						
Country	Yes	Yes	Yes	Yes	Yes	Yes

Note:
Dependent variable is defined as the average cosine similarity between dictionaries representing the object – politics – and the respective attributes – chaos and corruption. Robust standard-errors are clustered at the country level and reported in parentheses: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$

F.7 Minimum Threshold

Table 23: Impact of Regime Type on Association between Politics Object and Negative Attributes across Minimum Article Thresholds

	Chaos				Corruption			
	Minimum 300 (1)	Minimum 400 (2)	Minimum 500 (3)	Minimum 600 (4)	Minimum 300 (5)	Minimum 400 (6)	Minimum 500 (7)	Minimum 600 (8)
Independent Variables								
Xinhua	−11.88 (10.38)	−16.41 (9.93)	−17.13 (8.39)**	−15.33 (8.94)*	0.14 (11.51)	1.50 (10.55)	−0.51 (9.85)	1.16 (8.89)
Xinhua x Polity (IV)	0.61 (0.19)***	0.77 (0.15)***	0.68 (0.14)***	0.70 (0.14)***	0.53 (0.21)**	0.50 (0.23)**	0.34 (0.24)	0.42 (0.22)*
Control Variables								
Bilateral Treaty (Count)	7.67 (4.94)	7.92 (4.93)	11.27 (2.57)***	11.57 (2.57)***	6.10 (4.63)	7.43 (4.66)	9.94 (3.19)***	9.55 (2.99)***
Imports (USD)	−0.11 (0.37)	−0.22 (0.34)	−0.19 (0.33)	−0.21 (0.31)	0.02 (0.42)	−0.22 (0.41)	−0.28 (0.39)	−0.29 (0.38)
Country Terms (Freq)	−7.83 (2.08)***	−6.42 (1.79)***	−5.03 (1.65)***	−5.77 (2.08)***	−9.31 (1.93)***	−7.97 (2.11)***	−6.38 (2.01)***	−6.85 (2.06)***
Attribute Term (Freq)	−7.00 (1.60)***	−7.69 (1.36)***	−7.70 (1.31)***	−7.31 (1.67)***	−5.89 (1.43)***	−6.96 (1.63)***	−7.03 (1.50)***	−7.13 (1.48)***
Xinhua x GDP (log)	0.31 (0.68)	−0.03 (0.65)	−0.11 (0.47)	−0.21 (0.49)	0.66 (0.97)	0.57 (0.95)	0.62 (0.85)	0.13 (0.73)
Xinhua x GDPPC (log)	0.18 (1.20)	0.96 (1.18)	0.63 (0.87)	0.44 (0.91)	−0.40 (1.41)	−0.20 (1.36)	−0.34 (1.19)	−0.65 (1.07)
Xinhua x Inflation (%)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)*
Xinhua x Trade Balance (USD)	−0.02 (0.02)	−0.02 (0.02)	−0.02 (0.02)	−0.02 (0.02)	0.02 (0.04)	0.02 (0.04)	0.02 (0.03)	0.02 (0.03)
Xinhua x Conflict (Freq)	0.10 (0.27)	0.32 (0.26)	0.12 (0.26)	0.14 (0.26)	0.13 (0.27)	−0.05 (0.30)	−0.21 (0.33)	−0.23 (0.32)
Xinhua x Corruption Index	−0.41 (0.52)	−0.68 (0.53)	−0.30 (0.43)	−0.26 (0.43)	−0.13 (0.68)	−0.25 (0.70)	0.17 (0.60)	0.29 (0.60)
Xinhua x Country Terms (Freq)	−0.51 (1.48)	0.12 (1.47)	−0.65 (1.47)	−0.74 (1.47)	−1.89 (1.40)	−3.02 (1.27)**	−3.03 (1.18)**	−2.56 (1.21)**
Xinhua x Attribute Term (Freq)	0.46 (1.31)	−0.17 (1.36)	1.12 (1.20)	1.16 (1.24)	0.60 (1.19)	1.98 (1.14)*	2.45 (0.95)**	2.11 (1.16)*
Statistics								
Observations	140	132	128	125	140	132	128	125
Fixed effects								
Country	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note:
Dependent variable is defined as the average cosine similarity between dictionaries representing the object – politics – and the respective attributes – chaos and corruption. Robust standard-errors are clustered at the country level and reported in parentheses: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

F.8 Alternative Dependent Variable

Table 24: Impact of Regime Type on the Unweighted Average Similarity between Country labels and Negative Attributes

	Chaos				Corruption			
	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)	Model (7)	Model (8)
Independent Variables								
Xinhua	−12.00 (0.42)***	−45.30 (3.66)***			−14.00 (0.41)***	−25.14 (2.97)***		
CNA		−3.76 (0.63)***				−6.98 (0.55)***		
Polity (IV)	−0.55 (0.07)***	−0.22 (0.05)***			−0.45 (0.07)***	−0.18 (0.04)***		
Xinhua x Polity (IV)	0.28 (0.05)***	0.16 (0.04)***	0.30 (0.06)***	0.14 (0.04)***	0.22 (0.05)***	0.16 (0.04)***	0.24 (0.06)***	0.14 (0.04)***
Control Variables								
Country Terms (log)		−2.44 (0.27)***		−3.09 (0.34)***		−3.14 (0.16)***		−3.43 (0.27)***
Attribute Terms (Freq)		−8.62 (0.34)***				−7.15 (0.18)***		
Imports (USD)		−0.64 (0.20)***		−0.42 (0.24)*		−0.11 (0.17)		0.20 (0.23)
Trade Balance (USD)		0.00 (0.00)		0.00 (0.00)		0.00 (0.00)		0.00 (0.00)
GDP (log)		−0.29 (0.38)				−0.39 (0.31)		
GDPPC (log)		0.46 (0.36)				0.33 (0.34)		
Inflation (%)		0.00 (0.02)				−0.04 (0.02)**		
Bilateral Treaty (Count)		−2.15 (1.33)		−3.25 (1.40)**		−3.30 (1.02)***		−3.97 (1.23)***
Conflict (Freq)		0.48 (0.23)**				0.41 (0.18)**		
Corruption Index		−0.94 (0.23)***				−0.58 (0.19)***		
Xinhua x Country Terms (log)		0.29 (0.21)		0.49 (0.45)		0.68 (0.23)***		1.06 (0.30)***
Xinhua x Attribute Terms (Freq)		4.83 (0.43)***				1.83 (0.41)***		
Xinhua x Imports (USD)		−0.57 (0.27)**		−0.73 (0.42)*		−0.14 (0.27)		−0.57 (0.35)
Xinhua x Trade Balance (USD)		0.00 (0.00)		0.00 (0.00)		0.00 (0.00)		0.00 (0.00)
Xinhua x GDP (log)		0.84 (0.36)**		0.73 (0.51)		0.36 (0.36)		0.57 (0.47)
Xinhua x GDPPC (log)		−0.44 (0.29)		0.27 (0.40)		−0.34 (0.27)		0.29 (0.35)
Xinhua x Inflation (%)		0.02 (0.02)		0.02 (0.03)		0.04 (0.02)**		0.01 (0.02)
Xinhua x Bilateral Treaty (Count)		2.95 (1.75)*		4.01 (1.62)**		5.47 (1.20)***		6.44 (1.38)***
Xinhua x Conflict (Freq)		−0.40 (0.18)**		−0.07 (0.22)		−0.43 (0.14)***		−0.18 (0.14)
Xinhua x Corruption Index		0.43 (0.18)**		0.09 (0.23)		0.13 (0.16)		−0.15 (0.21)
Statistics								
Observations	15 828	13 483	15 828	13 483	15 828	13 483	15 828	13 483
Fixed effects								
FE	No	No	Yes	Yes	No	No	Yes	Yes

Note:
Dependent variable is defined as the unweighted average cosine similarity between dictionaries representing countries and the respective attribute. Robust standard-errors are clustered at the country level and reported in parentheses: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

Table 25: Impact of Regime Type on Unweighted Similarity between Politics Object and Negative Attributes

	Chaos				Corruption			
	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)	Model (7)	Model (8)
Independent Variables								
Xinhua	−12.18 (1.79)***	−36.84 (9.31)***	−12.18 (1.78)***	−27.34 (9.15)***	−11.40 (1.73)***	−19.41 (10.82)*	−11.29 (1.80)***	−9.34 (9.46)
Xinhua x Polity (IV)	0.62 (0.24)**	0.75 (0.18)***	0.62 (0.23)**	0.73 (0.14)***	0.80 (0.23)***	0.28 (0.21)	0.79 (0.24)***	0.34 (0.21)
Control Variables								
Country Terms (Freq)		−3.08 (1.24)**		−5.99 (1.54)***		−4.86 (1.84)**		−8.10 (1.98)***
Attribute Term (Freq)		−8.14 (1.05)***		−7.11 (1.21)***		−5.33 (1.59)***		−6.15 (1.50)***
Bilateral Treaty (Count)		3.55 (2.73)		6.74 (3.79)*		2.48 (3.13)		5.36 (4.23)
Imports (USD)		0.10 (0.30)		−0.04 (0.31)		−0.63 (0.45)		−0.11 (0.41)
Xinhua x Country Terms (Freq)		1.03 (1.64)		1.91 (1.47)		−1.75 (1.70)		−1.69 (1.19)
Xinhua x Attribute Term (Freq)		1.22 (1.59)		−0.59 (1.33)		2.94 (1.56)*		1.99 (1.15)*
Xinhua x GDP (log)		−0.70 (0.66)		−0.19 (0.60)		−0.01 (1.00)		0.22 (0.83)
Xinhua x GDPPC (log)		1.73 (1.47)		0.83 (1.15)		1.43 (1.71)		0.19 (1.21)
Xinhua x Inflation (%)		0.00 (0.00)		0.00 (0.00)		0.00 (0.00)**		0.00 (0.00)***
Xinhua x Trade Balance (USD)		−0.03 (0.02)		−0.02 (0.02)		0.00 (0.03)		0.01 (0.04)
Xinhua x Conflict (Freq)		0.34 (0.26)		0.38 (0.24)		−0.29 (0.37)		−0.15 (0.32)
Xinhua x Corruption Index		−0.50 (0.73)		−0.41 (0.54)		−0.61 (0.99)		−0.33 (0.66)
Statistics								
Observations	180	126	180	126	179	126	179	126
Fixed effects								
Country	No	No	Yes	Yes	No	No	Yes	Yes

Note:
Dependent variable is defined as the unweighted average cosine similarity between dictionaries representing the object – politics – and the respective attributes – chaos and corruption. Robust standard-errors are clustered at the country level and reported in parentheses: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$

Table 26: Impact of Regime Type on the Average Similarity between Country labels and Negative Attributes Over 1992-2010

	Chaos				Corruption			
	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)	Model (7)	Model (8)
Independent Variables								
Xinhua	2.00 (0.41)***	−20.65 (4.13)***			2.98 (0.36)***	−21.85 (2.66)***		
Polity (IV)	−0.38 (0.06)***	−0.18 (0.04)***			−0.28 (0.05)***	−0.16 (0.03)***		
Xinhua x Polity (IV)	0.01 (0.05)	0.12 (0.03)***	0.10 (0.06)*	0.10 (0.05)**	0.01 (0.04)	0.16 (0.03)***	0.11 (0.05)**	0.13 (0.04)***
Control Variables								
Country Terms (log)		−1.89 (0.24)***		−2.33 (0.32)***		−2.13 (0.14)***		−2.55 (0.28)***
Attribute Terms (Freq)		−7.21 (0.34)***				−7.19 (0.19)***		
Imports (USD)		−0.63 (0.20)***		−0.45 (0.28)		−0.11 (0.16)		−0.36 (0.27)
Trade Balance (USD)		0.00 (0.00)		0.00 (0.00)		0.00 (0.00)		0.00 (0.00)
GDP (log)		0.28 (0.34)				−0.05 (0.26)		
GDPPC (log)		0.34 (0.27)				0.22 (0.28)		
Inflation (%)		−0.01 (0.01)				−0.02 (0.01)**		
Bilateral Treaty (Count)		−0.35 (1.21)		−1.03 (1.03)		0.67 (1.02)		1.13 (1.25)
Conflict (Freq)		0.40 (0.16)**				0.54 (0.15)***		
Corruption Index		−0.67 (0.15)***				−0.53 (0.14)***		
Xinhua x Country Terms (log)		0.07 (0.25)		−0.11 (0.47)		−0.22 (0.25)		−0.05 (0.38)
Xinhua x Attribute Terms (Freq)		3.14 (0.45)***				3.11 (0.33)***		
Xinhua x Imports (USD)		−0.81 (0.23)***		−0.37 (0.34)		−0.22 (0.22)		−0.37 (0.31)
Xinhua x Trade Balance (USD)		0.00 (0.00)		0.00 (0.00)		0.00 (0.00)		0.00 (0.00)
Xinhua x GDP (log)		0.35 (0.31)		−0.09 (0.45)		0.03 (0.29)		0.02 (0.42)
Xinhua x GDPPC (log)		−0.42 (0.24)*		0.28 (0.33)		−0.26 (0.25)		0.27 (0.34)
Xinhua x Inflation (%)		0.02 (0.01)*		0.01 (0.02)		0.02 (0.01)**		0.01 (0.01)
Xinhua x Conflict (Freq)		−0.21 (0.15)		0.01 (0.20)		−0.53 (0.13)***		−0.30 (0.20)
Xinhua x Corruption Index		0.33 (0.14)**		−0.06 (0.18)		0.13 (0.14)		−0.10 (0.19)
Statistics								
Observations	19 919	12 223	19 919	12 223	19 919	12 223	19 919	12 223
Fixed effects								
FE	No	No	Yes	Yes	No	No	Yes	Yes

Note:
Dependent variable is defined as the weighted average cosine similarity between dictionaries representing countries and the respective attribute. Robust standard-errors are clustered at the country level and reported in parentheses: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$