



Diversionsary Media: Autocrat's Political Stabilization Tool During Political Unrest

Citation

Liu, Bonnie. 2023. Diversionsary Media: Autocrat's Political Stabilization Tool During Political Unrest. Bachelor's thesis, Harvard University Engineering and Applied Sciences.

Link

<https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37378270>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#)

Diversiory Media:
Autocrat's Political Stabilization Tool During Political Unrest

Bonnie Liu

Presented to the Department of Applied Mathematics
in partial fulfillment of the requirements
for a Bachelor of Arts Degree with Honors

Harvard College
Cambridge, Massachusetts
March 24, 2023

Diversionsary Media: Autocrat’s Political Stabilization Tool During Political Unrest

Bonnie Liu

March 24, 2023

Abstract

In face of negative political events that undermine political stability, does the government strategically ignore, distort, or distract? By examining the posting activity of the top 18 Chinese state-run media accounts on Weibo on days following unrest events, we find evidence for the third possibility: Chinese state-run media engage in strategic diversion to deflect public attention away from negative domestic political news. Specifically, it achieves so not through diversion with entertainment, but rather through diversion with political neutrality; namely, instead of catching the public eye with engaging or sensational nonpolitical content, the state-run media instead injects neutral and mundane political information that effectively dilutes the negativity without the need for factual distortion. For theory, we present a model of information manipulation and media management that relates the individual’s news consumption and protest decisions with the government’s media distortion and diversion. For empirics, we construct and perform topic classification on a novel dataset of all historical Weibo posts by top Chinese state-run media, and leverage political unrest data to empirically test various mechanisms of information manipulation and find that state-run media decrease the proportion of negative domestic news and increase the proportion of neutral domestic news, while roughly holding positive news constant. We also present a set of stylized facts that explore the heterogeneity of state-run media response to various types of unrest events, separated by actor, action, topic, and target type, as well as the heterogeneous responses between different types of media content.

Contents

1	Introduction and Conceptual Background	4
1.1	Diversionsary Tools	6
1.2	The Role of Media	7
1.3	Diversionsary Media	8
1.4	The Case of Chinese Government Media	11
1.5	Inattention	11
2	Model of Media Manipulation	13
2.1	Setup of Media Management and Information Transmission	14

2.2	Individual Problem	15
2.3	Government Problem	19
2.4	Static Game Nash Equilibrium with Naive Government	21
2.5	Sequential Game Nash Equilibrium with Sophisticated Government	21
2.6	A Visual Illustration	25
2.7	Operationalizing Distortion and Diversion through Weibo: A Toy Example	28
3	Data Collection and Processing	36
3.1	Measuring Political Unrest: GDELT Data	36
3.2	Measuring Government Media Activity: Weibo Data	40
3.2.1	Topic Classification	41
3.3	Instrumenting for Unrest: Weather Data	44
3.4	Variable Construction and Method of Aggregation	44
4	Empirical Methodology	46
4.1	Does the Chinese Government respond to Unrest by Engaging in Diversionary Media?	49
4.1.1	OLS Fixed Effects Regressions	49
4.1.2	Instrumental Variables Regression	51
4.2	Does government diversionary media effectively divert public attention?	54
4.2.1	OLS Fixed Effects Regressions	54
4.2.2	Instrumental Variables Regression	54
4.3	Potential Limitations	56
5	Results and Discussion	56
5.1	Does the Chinese Government respond to Unrest by Engaging in Diversionary Media?	58
5.1.1	By Broad Event Type	60
5.1.2	By Non-government Actor Action Type	61
5.1.3	By Target of Nongov Action	63
5.1.4	By Multiclass Topic Classification	64
5.1.5	By Destabilization Level	67
5.2	Does government diversionary media effectively divert public attention?	68
6	Conclusion	73
A	Appendix	75
B	Weibo Data	75
C	GDELT Data	77
D	Additional Regression Results	77

Acknowledgements

This thesis became a reality with the help of many people. I owe countless thanks to my advisors, friends, and family for their support for me throughout this process. In particular, I'd like to thank my advisors Professors David Yang and Edward Glaeser for their invaluable advice and patience, and for demonstrating to me the importance of economic research in understanding and improving our surrounding world. From the very first Econ 1011A lecture freshman fall to thesis meetings senior year, Prof. Glaeser sparked my interests in economics and continues to be an unwavering source of intellectual inspiration. Prof. Yang's insightful research on Chinese political economy inspired this thesis and his sound advice was equally crucial. I'm grateful for Giorgio Saponaro and Yihong Huang for their illuminating comments on the model. I thank Sean Anderson Ty and Boyuan Chen whose patience and insights on natural language processing and statistical inference were critical. I thank Professors Jesse Shapiro and Jaya Wen for encouraging the idea during the nascent stages, and Aakaash Rao for pushing me through. I thank my mentors Greg Bruich, Elden Elmanto, Kiran Gajwani, and Annie Park for their relentless encouragement for me in moments of self doubt throughout my college career. I thank my friends and Satoshi Yanaizu for supporting me through ups and downs. Lastly, I am forever indebted to my family, whose unconditional love and support for me have always been a ray of light through rain and shine.

1 Introduction and Conceptual Background

“A half truth is the worst of all lies, because it can be defended in partiality.” — Solon

“There are no whole truths, all truths are half truths.” — Alfred North Whitehead

Authoritarian governments have a hierarchy of arsenal in modifying their citizens’ behavior, from the the most overt and disagreeable to the most covert and palatable. The first tier is physical in nature, including violence and brute force. The second tier is economic in nature, including monetary and material rewards and punishments. The third tier is informational in nature, and relies on modifying the information available to the public. To nudge, tune, and herd the public’s behavior in desired directions, autocratic governments manipulate the composition of information in a way that is is optimally favorable to political stability.

This paper explores the third type of autocratic control through information manipulation and media management, and poses the following questions: In face of negative political events that undermine political stability, do autocratic governments strategically ignore, distort, or distract? Does it completely ignore negative news by not restricting media coverage akin to censorship, does it distort factual information to increase overall ppositivity, or does it distract the audience from negative news by inserting competing content? If it engages in distraction, does it do so by injecting eye-catching content unrelated to domestic politics, or by introducing mundane and neutral political facts to compete with negative political news?

To address these questions, we examine the case of China. Its unelected political leaders are known for government propaganda and censorship, making it an appropriate target for exploring whether authoritarian governments that heavily rely on media control do in fact engage in the diversion. I examine top Chinese state-run media’s Weibo activity following periods of heightened political unrest, and explore whether Chinese state-run media conduct diversion in the face of destabilizing events, what channels of diversion they deploy, and how successful their attempts are at directing the populace’s attention from unrest.

I estimate OLS fixed effects regressions and instrumental variables two stage least squares regressions and observe the following. First, Chinese state-run media engage in strategic diversion to deflect public attention away from negative domestic political news, by modifying its posting behavior. Specifically, it achieves so not through diversion with entertainment, but through

diversion with political neutrality: instead of crudely decreasing political coverage and increasing nonpolitical coverage, state-run media subtly avoid discussion of negative events and injecting neutral political information to diffuse attention and dilute negativity. This suggests the government conducts information management in a sophisticated way; it foregoes the obvious and detectable tactic of injecting eye-catching nonpolitical content, and instead pursues the nuanced maneuver of increasing neutrality of political news, while simultaneously decreasing negativity of political news. Second, this diversion is effective in simultaneously diverting audience attention away from negative domestic political content to prevent further destabilization, and attracting audience engagement with state-run media accounts overall to increase government control of the narrative. We additionally present a set of stylized facts that explores the heterogeneity across media topic categories and unrest event types.

The paper makes the following contributions. In terms of empirics, I construct a novel dataset of all historical Weibo posts by a large collection of top Chinese state-run media, and use supervised natural language processing methods for sentiment analysis topic classification. To the author’s knowledge, this is the most systematic large-scale analysis of the government information management in China. In terms of theory, I propose a model that unifies various forms of government information manipulation including diversion, falsification, propaganda, and censorship into a single coherent framework. Specifically, I consider media content as information signals for Bayesian updating, and government media management as attempting to influence the public’s posterior distribution of beliefs about the state of the world. I consider the distortion mechanism as changing the distribution’s first moment and the diversion mechanism as changing the distribution’s second moment.

This paper contributes to three main strands of extant literature, respectively about the government’s diversionary tools, the role of media, and limited attention in behavioral economics. The paper complements the literature on diversionary tools by considering media and entertainment as a diversionary tool, and contributes to the literature on the role of media by considering the diversionary nature of media as serving a political purpose despite its non-political content. It additionally combines insights from behavioral economics, micro-founding the government’s reaction function through individual optimization problems subject to Bayesian updating and limited attention.

1.1 Diversionary Tools

I contribute to the literature on autocratic governance through diversion by considering media and entertainment as a diversionary tool. In face of domestic turmoil, autocratic regimes often pacify the public by diverting attention away from politically contentious domestic concerns towards less destabilizing topics. They temporarily diffuse tension by deflecting scrutiny towards captivating events that carry less political risk. The diversionary toolkit comprises diversionary foreign policy (e.g. war), diversionary economic incentives (e.g. employment and transfers), and diversionary media which I propose.

Diversionary foreign policy

A first possibility is diversionary foreign policy. The government possesses a repertoire of diversionary foreign policy tactics; in face of economic declines or political pressures, governments sometimes resort to manufacturing foreign policy crises when they are unable to repress or reform domestically due to institutional constraints. To deflect attention from internal tensions to external tensions, the government instigates conflict abroad, which exerts both a cohesive effect via ingroup vs. outgroup dichotomy that reinforces ingroup solidarity through nationalist sentiment, as well as a scapegoating effect via blaming failed policies on foreign enemies (Oakes (2006)).

In the diversionary warfare theory, regimes threatened by domestic turmoil initiate military confrontations overseas to bolster political support (Sobek (2007) and Gelpi (1997)). For example, declining support for the prime minister's party in the U.K. is associated with an increased probability of the threat, display, or use of force (Morgan and Anderson (1999)); similarly, Israeli political protests led to an increase in the use of force by Israel (Sprecher and Karl Derouen (2002)). Cross-national studies also demonstrate that economic declines positively correlate with the probability of crisis escalation (James (1990)).

A softer tool compared to diversionary war is diversionary foreign rhetoric. Due to physical war's high cost and risk of escalation, leaders also employ less violent tactics like escalatory discourse instead of actual conflict to create a rally round the flag effect (Morgan and Bickers (1992); Kanat (2014)). Literature demonstrates that antagonistic rhetoric and aggressive foreign policy speeches can increase domestic support for a leader (Marra, Ostrom and Simon (1990); Brace and Hinckley (1993)). For instance, Carter (2020) shows that when unemployment and inflation rise, presidents

deploy "diversionary cheap talk" and criticize foreign nations (especially historical rivals) to boost approval ratings. Dictatorships are particularly disposed to engaging in this type of rhetorical strategy, due to their monopolization of political information and influence on popular opinion. Alrababa'h and Blaydes (2021) shows that following the outbreak of the civil war, the Syrian government disseminated conspiracy theories about Israel in government-controlled newspapers.

Diversionary Economic Incentives

A second and less controversial tool is diversionary economic incentive. Authoritarian regimes can offer various economic incentives to pacify unrest, and to divert attention away from underlying political issues. Economic incentives silence dissent from subpopulations facing poverty or discrimination by offering temporary economic relief, in order to avoid systematically resolving problems like poverty or inequality. For instance, in democratic regimes, politicians use government spending to increase chances of re-election, observed from an increase in transfers to voters (as a proportion of public spending) and increase in voter-targeted expenditures (Nordhaus (1975)), Akhmedov and Zhuravskaya (2004)). In autocratic regimes, Wen (2022) finds that the Chinese government uses government-subsidized state-owned enterprise (SOE) employment to promote social stability via job provision to male minorities.

One of my unique contributions to this literature is introducing media and entertainment as an additional diversionary tool. Distracting the public through social media is less risky and less controversial than warfare and less costly than economic rewards. Furthermore, while warfare generates heterogeneous reactions in the population, media consumption generates a more homogeneous effect as it is inherently less polarizing.

1.2 The Role of Media

The media plays an important role as the interface between government and the populace. Edmund Burke famously highlighted the importance of media, "There are three estates in Parliament but in the Reporters' Gallery there sits a Fourth Estate far more important." As the intermediary, mass media connects the public and the government in a bidirectional relationship.

Media plays two important roles of information acquisition (which concerns the receipt of information) and information control (which concerns the production and regulation of information).

The first role of information acquisition is bidirectional: the media informs the public about government policies and informs the government about aggregate preferences of the populace. The outcomes from information acquisition enable civilians to monitor government conduct and ensure government accountability. For instance, James M. Snyder and Strömberg (2008) finds that reduction in press coverage of politics adversely affects both voter information and politician action, manifested in lower federal spending and less effort from local congressmen. Gentzkow, Glaeser and Goldin (2004) examines the changes in the U.S. newspaper industry that occurred between 1870 and 1920 and relate them to the reduction of corruption in U.S. politics in the same period. However, since governments are held accountable only insofar as the public is informed about their policies, governments wish to develop favorable relationships with the media and exert political influence (Besley and Prat (2006)). Djankov et al. (2003) demonstrates how state ownership of the media is related to various measures of poor government performance.

The second role of information control is largely unidirectional: by exerting influence, whether partial or total, over the media, the government can engage in propaganda and censorship. Governments leverage media for propaganda (generating information favorable to the regime) and censorship (restricting information unfavorable to the regime). Through censorship, governments attempt to diffuse collective action threats and stymie popular uprisings (King, Pan and Roberts (2013)). More extremely, they may even impose digital sieges and restrict internet access during protest movements to quell unrest (Gohdes (2020), Swinhoe (2022)). Through propaganda, governments generate information that serves their agenda and promotes a biased perspective.

1.3 Diversionary Media

While propaganda and censorship have been extensively studied, the role of diversion has not been well documented. Beyond propaganda and censorship, governments also attempt to divert public attention during social unrest by changing the composition of media content. Diversionary media differs from the traditionally political use of media. While traditional media propaganda and censorship are inherently political in content, diversionary media can be entirely non-political. Since it's ideologically uncharged and uncontroversial, this mechanism is perceived to be less noticeable and abominable than propaganda and censorship. While propaganda involves forcefully indoctrinating the masses and inculcating political maxims, diversionary media involves subtly

incentivizing blissful ignorance.

Highly constraining forms of media management like censorship have their drawbacks. While intended to deter, censorship can instead intensify interest in censored topics, create opportunities for push-back, and signal government weakness. In fact, users who experience censorship are more likely to report being unfazed or angry about censorship than fearful or worried (Roberts (2018)), suggesting unintended consequences like risk of retaliation, undermined reputation, and increased (as opposed to decreased) curiosity in censored topics. Furthermore, once citizens are aware of censorship, in periods of upheaval, they may take the time to seek out restricted information and evade censorship, such as during the 2014 Hong Kong protests and the 2015 Tianjin explosion (Hobbs and Roberts (2018)). Since individuals are motivated to overcome frictions and seek information especially during crises, censorship is particularly ineffective when it is needed the most. Repression that encourages self censorship also creates information and surveillance problems for governments, as they often rely on input from the media and public sphere to identify local corruption and new pockets of dissent (Egorov, Guriev and Sonin (2009)).

By contrast, diversion mechanisms like flooding have two advantages; first, they effectively prey on individual's limited attention and cognitive capacity to indirectly achieve censorship; second, they are less conspicuous and objectionable, allowing the government plausible deniability. Diversionary flooding is the insertion of information aimed at distraction or confusion, which competes with information that reflects unfavorably on the regime by diluting it and distracting from it. While not forceful and not meant to persuade, these diversionary nudges create small inconveniences that reroute users to content that are more palatable to the regime, decreasing the mobilization capacity for opposition. As information is widespread and substitutable, small impediments like barriers to access and irrelevant spamming, not draconian punishments or sophisticated manipulation, can hugely affect citizen's consumption of political information. Through de-prioritizing and crowding out media deemed objectionable, governments have the power to selectively silence certain information channels detrimental to political stability, and nudge citizens away from activist circles, dangerous information, and focal points of coordination.

Diversionary media can manifest itself in reportage of captivating nonpolitical events, astroturfing by bots, or government mandated news articles. For instance, the Russian government uses armies of online bots and commentators to flood opposition hashtags and make it more difficult, but not

impossible, for people to find information on protests or opposition leaders (Goncharov (n.d.))

Concerning the first advantage of high efficacy, diversion preys on the public's limited attention (both time-wise in terms of short attention span and bandwidth-wise in terms of cognitive capacity) and lack of strong motivation to circumvent the diversion, thereby nudging them toward an information environment that is disconnected from politically sensitive realities. Since citizens become mired in the cacophony of media stimuli, their consumption of political information is highly sensitive to the costs of obtaining it. While distraction and misinformation can be discounted or avoided, they increase the burden on the public to separate signal from noise, authentic from embellished, reliable from fabricated, and pertinent from impertinent. This suggests massively different implications for information that is completely free and easy to obtain as compared to available but slightly difficult to access. Thus, even if censorship is porous and incomplete, flooding effectively diverts most people away from censored information.

Concerning the second benefit of low visibility: the dilution of core content by distracting content can be conceivably explained as algorithmic quirks, especially in the social media landscape where curated content is customized to individual preferences. Therefore, the government can avoid widespread use of repression, which often sparks popular backlash.

Primarily, diversion is instrumental in two scenarios: governments can use the media to distract people from both unpopular or problematic policies and overall negative state of affairs. For the former, governments may strategically time unpopular measures to coincide with newsworthy events that distract the media and the public. Durante and Zhuravskaya (2018) finds that Israeli attacks on Palestine are more likely to occur when U.S. news on the following day are dominated by important predictable events like the Olympic Games. For the latter, governments may strategically flood the media with irrelevant news following outbreak of destabilizing events to distract and appease the public. For instance, Barbera et al. (2020) finds that leaders worldwide strategically respond to unrest by simultaneously increasing their online messaging and shifting the topic from domestic to foreign policy issues. Besley and Dray (2023) finds that during the COVID-19 pandemic, countries without free media under-report COVID deaths by 3 to 6 deaths per million.

1.4 The Case of Chinese Government Media

As part of its media management toolkit, the Chinese government engages in diversionary flooding on media platforms. Roberts (2018) finds that most information management methods implemented by the Chinese government act not as a ban but as a tax on information: information is not impossible to access but accessing it requires extra time and money. The state is not interested in blocking sensitive information as much as redirecting the public’s attention from issues that could threaten the state’s legitimacy or authority. King, Pan and Roberts (2017) finds that the government fabricates social media posts by hiring bots to surreptitiously insert large amounts of pseudonymous writings into social media streams. By hiring paid commentators, governments can flood the internet discourse with their narrative, undermine the credibility of competing media, and distract citizens from events that reflect badly on them. As I find in the Weibo data, the government often distracts the public discourse with coverage of mundane details of Party meetings or with encouraging quotes. During times of greater unrest, the state run media may be more likely to release non-political and non-controversial “feel-good” content like nature documentaries or animal videos that divert the public’s attention away from political turmoil and elicit lighthearted satisfaction.

1.5 Inattention

In the behavioral economics literature, individuals simplify complex decisions, process subsets of information, and choose among alternatives with imperfect information. Specifically relevant to this paper is the impact of salience and limited attention on information acquisition and decision processes. Both theoretical models and field experiments demonstrate that salience (prominence of an attribute) significantly shapes preferences and behaviors (Bordalo, Gennaioli and Shleifer (2022)). Likewise for limited attention; in the rational inattention model, individuals allocate attention in a top-down way, using meta-cognition about what information should be acquired; they compare the costs of information acquisition and the benefits of decision improvement and deliberately ignore certain information in order to optimize their attentional resources (Sims (2003)). In behavioral inattention, individuals allocate attention in a bottom-up way, and unintentionally overlook information that is relevant to decision-making, resulting in suboptimal outcomes. Various

factors contribute to this phenomenon, including the information’s salience of presentation and human limitations in attentional capacity or cognitive biases under information overload.

Examples of inattention are abundant: in the public economics domain, Chetty, Looney and Kroft (2009) finds that consumers are inattentive to indirect state taxes that are not transparently factored into prices. In the financial economics domain, limited attention among investors induces under-reaction to newly released information, contributing to anomalies like momentum. Investors are inattentive to and underreact to various kinds of information, including 1) earnings announcement (as inattention decreases the immediate stock price response and increases the future predictability of returns) (Bernard and Thomas (1989)), 2) indirect economic links between supplier and customer companies (Cohen and Frazzini (2008)), and 3) news releases during times of greater distraction (e.g. earnings announcements on Fridays features less immediate and more delayed response compared to average (Dellavigna and Pollet (2009))).

This paper contributes to this literature by considering application of limited attention and salience to media consumption in the political economy realm. In the media context, more salient (eye-catching) content is more likely to attract the audience’s attention. Literature has documented the interplay between attention and media consumption. The effectiveness of information transmission through mass media depends on the presence of other attention grabbing (whether amusing, sensational, fear inducing, etc.) content that may crowd out important information about the state of the world. For instance, Eisensee and Strömberg (2007) find evidence supporting that U.S. government decisions on natural disaster relief depend on the news coverage of disasters, and the presence of other newsworthy events like the Olympic Games crowds out disaster coverage. Additionally, individuals face an attention constraint and can only process a fixed amount of information per unit time; therefore, they selectively allocate attention in a bottom-up way, as attention is not predicated upon meta-level decision, but depends on presence of salient features. The use of diversionary media by the Chinese authoritarian government, if existent and successful, would provide a real-life and politically important evidence of behavioral inattention in media consumption that substantiates these theoretical literature. Furthermore, this paper’s model of media manipulation combines insights from behavioral economics, micro-founding the government’s reaction function through individual optimization problems subject to Bayesian updating and limited attention.

The rest of the paper proceeds as follows. Section 2 presents a parsimonious microfounded model of diversionary media. Section 3 discusses the data collection and processing. Section 4 discusses the empirical design and regression specifications. Section 5 overviews select empirical results. Finally, section 6 concludes by discussing implications of our findings and directions for further research.

2 Model of Media Manipulation

I propose a partial equilibrium model of both the demand and supply sides of media and protest to illustrate the key dynamics in a regime with government control of media. I focus on the market of media content: the public consumes content while the government produces content. I connect the government problem and the individual problem via media manipulation and protest, and in so doing microfounds diversionary media in the individual optimization problem. On the government side, the government maximizes political stability determined by both unrest and public trust in the government; the government's constraint comes from costs of media manipulation like diminished trust and increase perceived severity of the issue. On the individual side, the individual minimizes the error between action under full information (i.e. if the individual observes true state instead of signal) and action under only partial information (i.e. if the individual only observes signal); the individual's constraint comes from limited information. To make our model parsimonious in assumptions and generalizable, I assume both demand and supply of media (both the government and the public) are rational, in order to avoid level-k theory complications in which actors view others as less rational and sophisticated than themselves.

I incorporate diversion, falsification, censorship, and propaganda into a unified model. Specifically, I do so by modeling the state of the world and the noise manufactured by the government as normal distributions, which gives the government two channels for media manipulation/information management: 1) changing the first moment (level) of the signal, i.e. how positive or negative the signal is, and 2) changing the second moment (noise) of the signal, how much variance the signal contains. I consider falsification, censorship, and propaganda as all shifting the first moment of the distribution of news content, and consider diversion as shifting the second moment of the distribution.

The order of events are displayed in Figure 1: Individuals take the signal that incorporates diversion and falsification as exogenous, and choose the optimal amount of protest; next, the government observes the level of protest from the individual problem and chooses the optimal response of falsification and diversion.

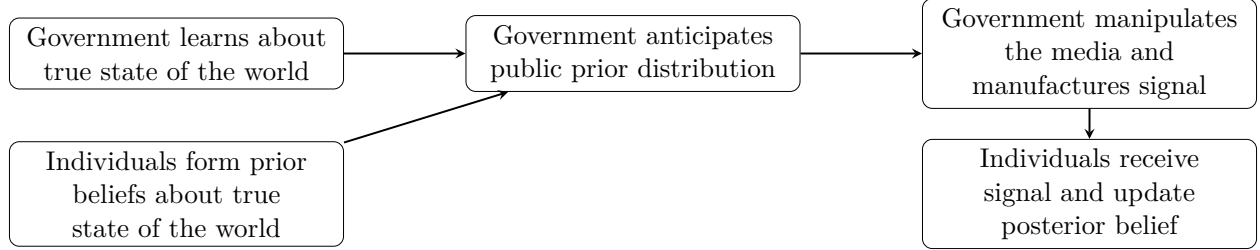


Figure 1: Sequence of Events in Model Setup

2.1 Setup of Media Management and Information Transmission

The government understands that the individual has a prior distribution and that the individual Bayesian updates to form a posterior distribution of beliefs; therefore, the government’s objective is to influence the posterior distribution.

The government knows the true value x but strategically manufactures a noise d and combines truth and noise into a signal s sent to the public. The signal equals the true value plus noise: $s = x + d$. The true value x and noise d are independent and both distributed normal. Since x and d are both distributed normal, $s = x + d$ is a mixture of normals. Note that x is measure of “negativity” of the state of affairs; therefore, the higher the x , the worse the state of the world.

$$x \sim \mathcal{N}(\mu_x, \sigma_x^2)$$

$$d \sim \mathcal{N}(\mu_d, \sigma_d^2)$$

$$x \perp\!\!\!\perp d$$

The mean and variance parameters that capture the first and second moments of the distribution take on the following interpretation: μ_d is the government’s misrepresentation of the factual state of the world; a negative μ_d indicates that the government decreases the first moment (i.e. make the news more positive). σ_d is the government’s injection of diversionary content; a positive σ_d indicates

that the government increases the second moment (i.e. make the news have higher variance). In this framework, diversion changes the second moment of the distribution, while falsification/fabrication changes the first moment of the distribution.

Importantly, distortion and diversion achieve the same result through different channels: distortion creates a wedge between the true state and the signal, whereas diversion creates a wedge between the signal and the inference. Under distortion (manipulation of first moment), the government sends a signal that is increasingly biased in the negative direction. Under diversion (manipulation of second moment), the government sends a signal that is truthful in expectation (level) but inflated with noise; due to the noisiness of the signal, individuals shrink the signal towards the cognitive default. The diversion channel bears resemblance to cognitive uncertainty (as in Enke and Graeber (2019)) except being fully rational. Importantly, noise and bias are linked; namely, when people are cognitively noisy, they revert more to a cognitive default, which introduces systematic bias. Mathematically, the second moment of the signal is relevant for inferring the first moment.

The mean and variance can be manipulated in several ways. To change the mean μ_d of politically relevant news, the government can engage in propaganda, censorship, embellishment, and fabrication. For instance, the government can limit the news coverage of negative domestic political news (i.e. censorship); shift the sentiment distribution of domestic political posts from negative to positive, either through making negative events sound more positive (i.e. falsification), or through substituting posts on negative events with posts on positive events (sentiment distortion); shift the location distribution of political news from domestic to foreign (i.e. location diversion); shift the topic distribution of domestic news from political to nonpolitical (i.e. topic diversion). The conceptual framework explains each mechanism in greater detail. To change σ_d of domestic political news, the government can decrease the number of domestic political posts, and instead inject other categories of information, including foreign events or nonpolitical topics.

2.2 Individual Problem

The country has a unit mass of individuals each with prior and posterior beliefs about politically relevant issues. Individuals update their prior beliefs in the classical Bayesian manner after receiving signal from their media intake to form a posterior distribution. Each individual makes the binary decision of whether or not to protest. Aggregated across individuals, the level of protest becomes

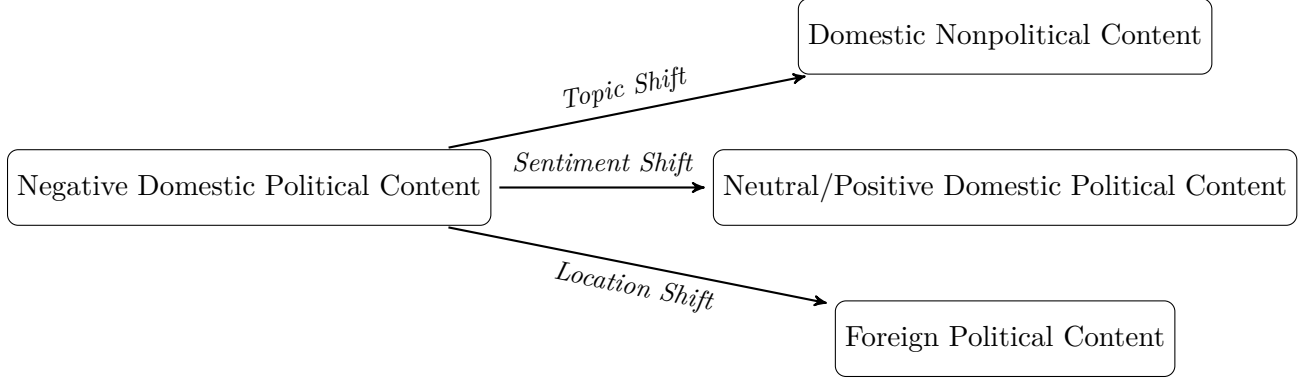


Figure 2: Flowchart of Distortion and Diversion Mechanisms

a continuous fraction (taking on a probabilistic interpretation in case of unit mass). I work with the aggregate continuous case for sake of simplicity.

Individuals receive signal s but do not observe the true value x . The individual subsequently updates their prior belief. The agent chooses a level of p that potentially differs from the fully informed action $p^{informed} = f(x)$, where x is the true state of the world. The agent wants to minimize the distance between his partially informed action given the signal (p) and the fully informed action ($p^{informed}$), where the action is how much to engage in protests. Setting $f(x) = x$, the agent solves

$$\min_p u(p) = \frac{1}{2} \mathbb{E}[(p - x)^2 | s]$$

Expanding and differentiating with respect to p , I obtain

$$\begin{aligned} \min_p \frac{1}{2} (p^2 - 2p\mathbb{E}[x|s] + \mathbb{E}[x^2|s]) \\ p^* = \mathbb{E}[x|s] \end{aligned}$$

The process of identifying the rational action $p^{informed}$ is imperfect because agents do not observe the true value; thus, the inferred posterior is a convex combination of the signal and the default. For a Bayesian agent, this creates a Gaussian signal extraction problem. Let y denote the posterior distribution of x given signal s , such that $y = \mathbb{P}(x | s)$.

$$y = \mathbb{P}(x | s) \sim \mathcal{N}(\lambda s + (1 - \lambda)\mu_x, (1 - \lambda)\sigma_x^2) \text{ for shrinkage factor } \lambda = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_d^2} \in [0, 1]$$

The above summarizes the outcome of the government using both levers of distortion and diversion; I now isolate the effects of each lever. When there is only distortion and no diversion ($\sigma_d^2 = 0$), there is no shrinkage towards the prior default ($\lambda = 1$), and the posterior distribution reduces to $Y \sim \mathcal{N}(s, 0)$. When there is only diversion and no distortion ($\mu_d^2 = 0$), then the signal's value is just the prior default ($s = \mu_x$), and the posterior distribution reduces to $Y \sim \mathcal{N}(\mu_x, (1 - \lambda)\sigma_x^2)$.

Since $\text{Cov}(x - \lambda s, s) = 0$, $(x - \lambda s, s)$ is Multivariate Normal random variable with uncorrelated (and hence independent) components. in particular, due to independence of components

$$\mathbb{E}[x - \lambda s \mid s] = \mathbb{E}[x - \lambda s] = (1 - \lambda)\mu_x - \lambda\mu_d = \frac{\sigma_d^2}{\sigma_d^2 + \sigma_x^2}\mu_x - \frac{\sigma_x^2}{\sigma_x^2 + \sigma_d^2}\mu_d$$

Proposition 1. *The individual's optimal protest level is given by*

$$p^* = \mathbb{E}[x \mid s] = \frac{\sigma_d^2}{\sigma_d^2 + \sigma_x^2}\mu_x - \frac{\sigma_x^2}{\sigma_x^2 + \sigma_d^2}\mu_d + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_d^2}s$$

Proposition 2. *We verify comparative statics as desired: first, optimal protest is decreasing in distortion; second, optimal protest is decreasing in diversion if the signal is high (bad) enough (recall s measures amount of “badness”), and increasing in diversion if the signal is low (good) enough; i.e. diversion decreases protest if the realized signal is sufficiently bad, but counterproductively increases protest if signal is sufficiently positive.*

$$\begin{aligned} \frac{\partial p^*}{\partial \mu_d} &= -\frac{\sigma_x^2}{\sigma_x^2 + \sigma_d^2} \in [-1, 0] \\ \frac{\partial p^*}{\partial \sigma_d^2} &= -(s - \mu_d - \mu_x) \frac{\sigma_x^2}{(\sigma_x^2 + \sigma_d^2)^2} \\ \frac{\partial p^*}{\partial \sigma_d^2} &\begin{cases} > 0 & \text{if } \mu_x + \mu_d > s \text{ i.e. signal is less than sum of default values of prior and signal} \\ < 0 & \text{if } \mu_x + \mu_d < s \text{ i.e. signal is greater than sum of default values of prior and signal} \end{cases} \end{aligned}$$

where the second comparative static comes from

$$\frac{\partial p^*}{\partial \sigma_d^2} = \frac{\partial p^*}{\partial \lambda} \frac{\partial \lambda}{\partial \sigma_d^2}$$

$$\frac{\partial p^*}{\partial \lambda} = s - \mu_d - \mu_s = \begin{cases} < 0 & \text{if } \mu_x + \mu_d > s \\ > 0 & \text{if } \mu_x + \mu_d < s \end{cases}$$

$$\frac{\partial \lambda}{\partial \sigma_d^2} = -\frac{\sigma_x^2}{(\sigma_x^2 + \sigma_d^2)^2} < 0$$

Proposition 3. *The positive cross comparative static is given by the below. Since each individual comparative static is negative, this positive cross comparative static indicates diversion and distortion are substitutes for influencing protest.*

$$\frac{\partial p^*}{\partial \mu_d \partial \sigma_d^2} = \frac{\sigma_x^2}{(\sigma_x^2 + \sigma_d^2)^2} > 0$$

The comparative statics and cross partial confirm our intuition. The more the government fabricates noise, the more the population is content, due to either being satisfied from the positiveness of news (from positive distortion, higher μ_d) or being distracted by politically irrelevant news (from higher diversion, higher σ_d). For the first comparative static $\frac{\partial p^*}{\partial \mu_d}$, when the government does not engage diversion (i.e. $\sigma_d^2 = 0$), then $\frac{\partial p^*}{\partial \mu_d} = -1$, hence distortion is highly effective and decreases individual's protest level one-to-one. Meanwhile, when the government engages in a lot of diversion (i.e. $\sigma_d^2 \rightarrow \infty$), then $\frac{\partial p^*}{\partial \mu_d} \rightarrow 0$, hence distortion becomes less effective at decreasing protest as diversion increases.

For the second comparative static $\frac{\partial p^*}{\partial \sigma_d^2}$, I observe that the effectiveness of diversion (changing second moment of signal) in deterring protest depends on the mean of the prior and distortion (first moment of signal). Since x and s measures the amount of negativeness in the world (e.g. unrest elsewhere), when the signal is higher (i.e. worse) than the default prior plus the expected distortion, then diversion decreases individual's protest level, as increasing variance of the noise flattens the distribution of the posterior, which induces more density to the right of the protest threshold, therefore decreasing level of protests. Vice versa, when the signal is lower (i.e. better) than the default prior plus expected distortion, then diversion increases optimal protest, as flattening the

posterior distribution via increasing variance counterproductively induces more mass to the right of the protest cutoff threshold in this case.

Meanwhile, the positive cross comparative static $\frac{\partial p^*}{\partial \mu_d \partial \sigma_d^2}$ indicates that distortion and diversion are substitutes in decreasing individual's optimized protest level. This confirms our conceptual model of the government having two simultaneous levers for media manipulation, and can choose to pursue either while accounting for the different relative costs on public trust and reputation.

2.3 Government Problem

The government takes the optimal level of protest from the individual problem as input, and maximizes political stability over the endogenous variables μ_d and σ_d^2 . I model a simple government whose utility function only consists of political stability. Its utility function is given by

$$U^{Gov} = U(t, p) = t - p \text{ where } \begin{cases} U'_t > 0 \\ U'_p < 0 \end{cases}$$

where t is a cost function (economically, t could represent trust the public has in the government), and p is the level of protest activity from individuals. Political stability is increasing in trust in the government $U'_t > 0$, and decreasing in amount of protest $U'_p < 0$.

Mathematically, cost t imposes a constraint on government media manipulation by introducing a cost of manipulation. This cost of trust can be conceptually understood as a psychological cost of diminished public trust in the government (Hensel et al. (2020)), or cost of higher perceived severity, or even a concrete material cost of creating the media content.

For the government utility function, protest level p comes from the individual problem, while cost t does not. The cost t only enters into the government problem as opposed to the individual problem, unrest p is directly determined by the individual optimization problem and is therefore a well specified function for the government.

Let trust

$$t = v(\mu_d, \sigma_d) = -\mu_d^2 - \delta \sigma_d^2 \text{ where } v'(\sigma_d^2) < 0, \text{ and } \begin{cases} v'(\mu_d) < 0 \text{ when } \mu_d > 0 \\ v'(\mu_d) > 0 \text{ when } \mu_d < 0 \end{cases}$$

Trust takes on the additively separable functional form in its arguments: the level (mean μ_d) and spread (standard deviation σ_d) of the noise. Level of trust in the government suffers the adverse consequences of falsification and diversion. If we interpret t as the cost of media manipulation on trust, t is maximized when $\mu_d = 0$ and $\sigma_d = 0$ (i.e. the most trustworthy government is one that engages in neither distortion nor diversion). Cost is quadratical in distortion, as the probability of the public detecting the government's lying increases at a positive rate. This dynamic between government media intervention and public trust in the government can also be modeled intertemporally, where each period's media intervention erodes public trust in the government in the next period. However, a static one period setup suffices to deliver the qualitative predictions I hypothesize conceptually and observe empirically.

The scalar δ motivates the government's decision between changing μ_d or σ_d . When $\delta < 1$, distortion is more costly for trust than diversion; whereas when $\delta > 1$, diversion is more costly. Conceptually, changing the first moment is more noticeable and potentially repellent than changing the second moment. Namely, engaging in information falsification, repressive censorship, or nationalistic propaganda is much more likely to result in detection by the public than engaging in softer methods like distraction through flooding; therefore, distortion (artificially increasing the mean of the signal) is more likely to erode trust in the government than diversion (artificially increasing the variance of the signal). We set $\delta = 1$ for simplicity.

As encapsulated in the above specification, government manipulation of media, including both changes in the first moment (i.e. distortion/fabrication) and changes in the second moment (i.e. diversion and distraction), simultaneously generates two competing effects. First, media manipulation reduces civilian trust in the government, as the public becomes cognizant of the government's ulterior motive of appeasing the public by injecting positive or distracting content into their media feed. Second, media manipulation decreases the amount of unrest potential, since a greater fraction of the population is distracted.

Therefore, the government maximizes its utility function $U^{Gov} = U(t(\mu_d, \sigma_d), p(\mu_d, \sigma_d))$ and solves for μ_d^* and σ_d^* , where μ_d^* represents the optimal level of distortion and σ_d^* captures the optimal level of diversion. Below we discuss two different Nash Equilibrium resulting from two different games: a static game and a sequential game.

2.4 Static Game Nash Equilibrium with Naive Government

In the static version of the game, the government treats the protest level as exogenous and does not anticipate that protest responds to media manipulation and does not treat p as function of μ_d and σ_d . In other words, the static game models a naive government who solves:

$$\max_{\mu_d, \sigma_d} U_{\mu_d, \sigma_d}^{Gov}(\mu_d, \sigma_d; p) = -\mu_d^2 - \sigma_d^2 - p$$

The solution is immediate: if the government believes protestors engage in unrest regardless of media content, media information management portends no returns and only delivers the downside the diminished trust. In this eventuality, the government does not engage in distortion or diversion and instead reports truthfully, with $\mu_d^* = 0$ and $\sigma_d^{2*} = 0$. Next substitute the government's optimal response μ_d^* and σ_d^{2*} into the individual's optimal response p^* .

Proposition 4. *When the government does not internalize the individual response function, the equilibrium is a trivial one where the government engages in neither diversion nor distortion, and the individual protests to a commensurate degree as the signal. The Nash Equilibrium is characterized by best response functions*

$$p^*(\mu_d^*, \sigma_d^*) = s \tag{1}$$

$$\mu_d^* = 0 \tag{2}$$

$$\sigma_d^{2*} = 0 \tag{3}$$

2.5 Sequential Game Nash Equilibrium with Sophisticated Government

In the sequential version of the game, the government internalizes $p^*(\mu_d, \sigma_d^2)$ and solves

$$\begin{aligned} \max_{\mu_d, \sigma_d} U_{\mu_d, \sigma_d}^{Gov}(\mu_d, \sigma_d, p(\mu_d, \sigma_d)) &= -\mu_d^2 - \sigma_d^2 - p(\mu_d, \sigma_d) \\ &= -\mu_d^2 - \sigma_d^2 - ((1 - \lambda)\mu_x - \lambda\mu_d + \lambda s) \text{ where } \lambda = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_d^2} \end{aligned}$$

Proposition 5. *Government utility obtains its maximum at the following optimal values*

$$\mu_d^* = \frac{1}{2} \frac{\sigma_x^2}{\sigma_x^2 + \sigma_d^2} \quad (4)$$

$$\sigma_d^{2*} = \begin{cases} (\mu_x + \mu_d - x - d - 1)\sigma_x^2 & \text{when } (x - \mu_x) + (d - \mu_d) < -1 \\ 0 & \text{when } (x - \mu_x) + (d - \mu_d) > -1 \end{cases} \quad (5)$$

Optimal diversion σ_d^{2*} has an interior solution when $x < \mu_x + \mu_d - 1 - d$, and has a corner solution at 0 when $x > \mu_x + \mu_d - 1 - d$. This result comes from the F.O.C. which arranges to $\frac{\sigma_x^2}{(\sigma_x^2 + \sigma_d^{2*})} = \frac{1}{\mu_x + \mu_d - s}$. Since the left hand side is between 0 and 1, the FOC only reaches a positive solution when $x < \mu_x + \mu_d - 1 - d$. Otherwise, $\frac{\partial U^{Gov}}{\partial \sigma_d^2} < 0$ for all positive σ_d^2 , indicating that any positive amount of diversion will decrease government utility; since diversion is non-negative, we reach the corner solution where optimal diversion is zero.

The intuition behind this case-wise result is as follows. In case 1, when the true value x is lower (better) than threshold, the default prior is lower than the posterior ($x < \mathbb{P}(x|s)$); thus, increasing noise σ_d shifts the posterior belief downward towards the default prior (recall that downward shift means lower level of “badness”), thereby improving people’s belief about the state of the world and reducing level of protests. In case 2, when true value x is higher (worse) than threshold, the default prior is higher than the posterior ($x > \mathbb{P}(x|s)$); thus, increasing noise σ_d shifts the posterior belief upward towards the default prior, thereby increasing people’s posterior beliefs about the existing level of unrest and increasing protests. Since diversion increases unrest in this case, the government does not engage in diversion.

Alternatively, consider the sum of differences between realizations and expectations $(x - \mu_x) + (d - \mu_d)$. When this quantity is higher in the positive direction, it indicates that the random variables x and d realized at higher values than their expectations, where higher values indicate more problematic state of the world. Thus, $(x - \mu_x) + (d - \mu_d) > -1$ translates into a bad realization of both the true value and the noise in the first moment, in this case, the second moment can help to improve posteriors; hence σ_d^2 is positive. By contrast, $(x - \mu_x) + (d - \mu_d) < -1$ translates into a good realization; thus, since the first moment of the posterior is already sufficient good, diversion is no longer necessary; hence σ_d^2 is zero.

Note that since signal is distributed $s \sim \mathcal{N}(\mu_x + \mu_d, \sigma_x^2 + \sigma_d^2)$, the threshold $s = x + d = \mu_x + \mu_d - 1$ is lower than the mean. Therefore, the fact that the cutoff rule is lower means the government is tolerating certain bad states of the world and only deploying diversion when the state of the world is negative enough.

Proposition 6. *The Stackelberg equilibrium is characterized by the below system of optimal response functions. When both players are playing this NE strategy, neither party has an incentive to deviate.*

$$\mu_d^* = \frac{1}{2} \frac{\sigma_x^2}{\sigma_x^2 + \sigma_d^{2*}} \quad (6)$$

$$\sigma_d^{2*} = (\mu_x + \mu_d - s - 1)\sigma_x^2 \text{ for } x < \mu_x + \mu_d^* - 1 - d \quad (7)$$

$$p^* = \frac{\sigma_d^{2*}}{\sigma_d^{2*} + \sigma_x^2} \mu_x - \frac{\sigma_x^2}{\sigma_x^2 + \sigma_d^{2*}} \mu_d^* + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_d^{2*}} s \quad (8)$$

Next, we solve for $\mu_d^*, \sigma_d^{2*}, p^*$ in terms of exogenous variables and implicitly differentiate with respect to x , μ_x , and σ_x^2 to obtain comparative statics, for μ_d and σ_d^{2*} .

Since

$$2\sigma_d^{2*} + \sigma_x^2(2 - \mu_x + s) \begin{cases} > 0 \text{ when } x > -2\frac{\sigma_d^2}{\sigma_x^2} + \mu_x - 2 - d \text{ which we call A for convenience} \\ < 0 \text{ when } x < -2\frac{\sigma_d^2}{\sigma_x^2} + \mu_x - 2 - d \end{cases}$$

Proposition 7. *The government's optimal distortion μ_d^* and diversion σ_d^{2*} are characterized by the following comparative statics with respect to x , μ_x , and σ_x^2 .*

$$\frac{\partial \sigma_d^{2*}}{\partial x} = \frac{\overbrace{-(\sigma_x^4 + \sigma_x^2 \sigma_d^{2*})}^{<0}}{2\sigma_d^{2*} + \sigma_x^2(2 - \mu_x + s)} \begin{cases} > 0 \text{ when } x < A \\ < 0 \text{ when } x > A \end{cases} \quad (9)$$

$$\frac{\partial \sigma_d^{2*}}{\partial \mu_x} = \frac{\sigma_x^4 + \sigma_x^2 \sigma_d^{2*}}{2\sigma_d^{2*} + \sigma_x^2(2 - \mu_x + s)} \begin{cases} > 0 \text{ when } x > A \\ < 0 \text{ when } x < A \end{cases} \quad (10)$$

$$\frac{\partial \sigma_d^{2*}}{\partial \sigma_x^2} = \frac{2\sigma_x^2(\mu_x - s - \frac{1}{2}) - \sigma_d^2(2 - \mu_x + s)}{2\sigma_d^{2*} + \sigma_x^2(2 - \mu_x + s)} \quad (11)$$

$$\frac{\partial \mu_d^*}{\partial x} = -\frac{1}{2} \frac{\sigma_x^2}{(\sigma_d^{2*} + \sigma_x^2)^2} \frac{\partial \sigma_d^{2*}}{\partial x} \text{ which has opposite sign as } \frac{\partial \sigma_d^{2*}}{\partial x} \quad (12)$$

$$\frac{\partial \mu_d^*}{\partial \mu_x} = -\frac{1}{2} \frac{\sigma_x^2}{(\sigma_d^{2*} + \sigma_x^2)^2} \frac{\partial \sigma_d^{2*}}{\partial \mu_x} \text{ which has opposite sign as } \frac{\partial \sigma_d^{2*}}{\partial \mu_x} \quad (13)$$

$$\frac{\partial \mu_d^*}{\partial \sigma_x^2} = \frac{1}{2} \frac{\sigma_d^{2*} - \sigma_x^2 \frac{\partial \sigma_d^{2*}}{\partial \sigma_x^2}}{(\sigma_d^{2*} + \sigma_x^2)^2} \quad (14)$$

These comparative statics show how the optimal distortion and diversion change with respect to the true state of the world, as well as mean and variance parameters of its underlying distribution. A higher x indicates a worse draw from the underlying probability distribution. A higher μ_x indicates that overall state of affairs is worse in expectation, even before the random variable realizes a value; this could imply a uniform upward shift of the entire distribution, or a shift in probability mass of the distribution from left end to right end. For instance, a higher μ_x could translate into overall worse economic conditions. A higher σ_x^2 indicates that state of affairs is high variance and each realization can deviate far from the mean. For instance, a higher σ_x^2 could translate into greater uncertainty about the future economic or political landscape, such as leading up to party leadership change or during a trade war. Therefore, the comparative statics indicate not only how optimal media policy responds to a particular realization, but also to underlying fundamental conditions.

Equation 12 ($\frac{\partial \mu_d^*}{\partial x}$) shows that when the true value is below (better than) a certain threshold (dictated by μ_x and σ_x^2), optimal state-run media distortion decreases as the true state of the world worsens, i.e. as the true state of the world worsens, the optimal distortion is lower; to the contrary, if the true value is above (worse than) certain threshold, optimal media distortion increases as the true state of the world worsens; namely, it's optimal to increase diversion when the state of affairs worsens. Together, these two cases suggest that the response of optimal distortion is dependent on the true level of unrest x . Colloquially, when the draw is good enough, the state would optimally distort less as the draw incrementally gets worse; whereas when the draw is bad enough, the state would optimally distort more if the draw incrementally gets worse.

Equation 9 ($\frac{\partial \sigma_d^{2*}}{\partial x}$) shows that when true unrest is above (worse than) certain threshold, optimal state-run media diversion is decreasing in the true unrest; vice versa, when true unrest is better

than cutoff, diversion increases as the state of affairs deteriorates. Taken together, the above comparative statics suggest that distortion and diversion are substitutes for the government.

2.6 A Visual Illustration

We now present a simplified visual illustration of the above model. Figure 3 displays how diversion changes the individual's inference from the government's signal, via graphing the two limit scenarios: when $\sigma_d = 0$ and when $\sigma_d = \infty$. With no diversion, the inference is precisely equal to the signal. With infinite diversion, the individual recognizes the noise and infinitely defaults their posterior to the prior, causing the inference to be unresponsive to the signal. Figure 4 shows the state of the world and the corresponding diversion. When true unrest is low (x is low), the government does not divert, whereas when true unrest is high, the government injects diversionary content and chooses positive σ_d^2 .

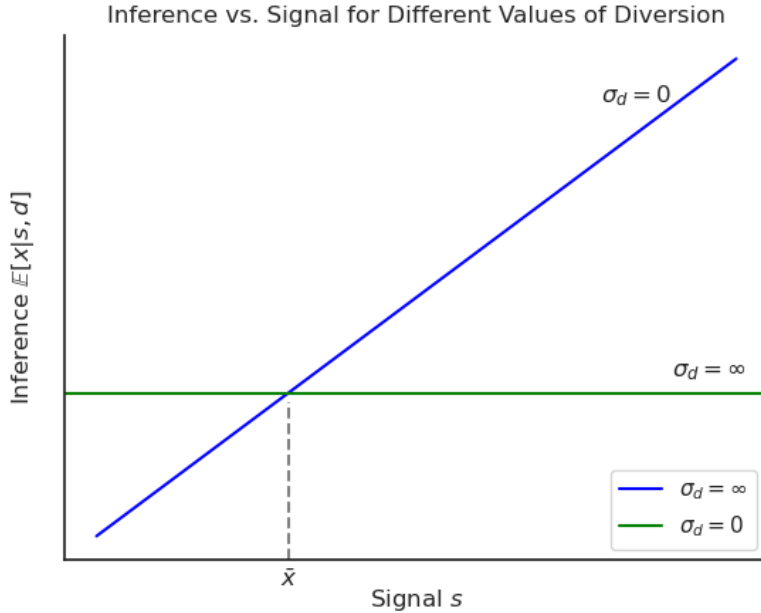


Figure 3: Ranges of Diversion and Corresponding Inference vs. Signal Function

Figures 5 and 6 illustrate how mechanistically distortion and diversion impact the signal and inference. Under distortion (of first moment), the government sends a signal that is increasingly biased downwards as unrest increases. Under diversion (of second moment), the government sends a signal that is truthful in expectation (level) but inflated with noise; due to the noisiness of the

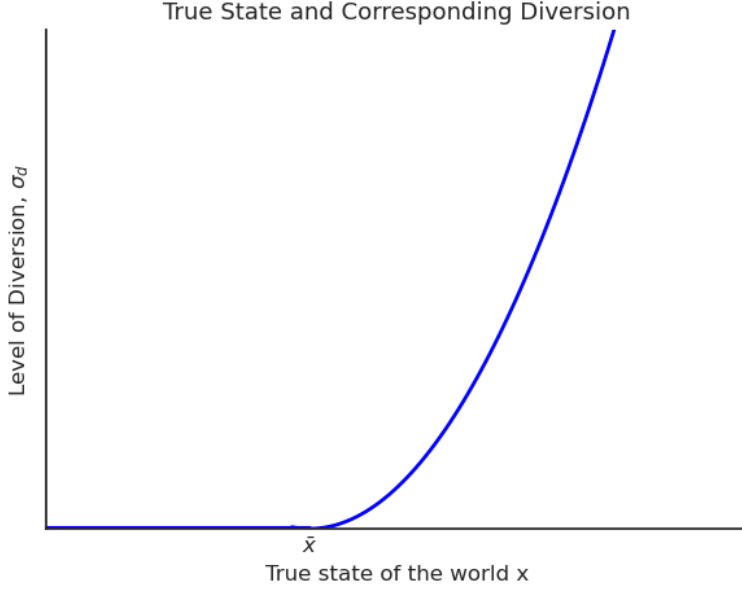


Figure 4: True State of the World and Corresponding Diversion

signal, individuals shrink the signal towards the cognitive default. Importantly, we observe that distortion and diversion achieve the same result through different channels: distortion creates a wedge between the true state and the signal, whereas diversion creates a wedge between the signal and the inference.

As shown in graph 7, the government seeks to increase the probability mass under the protest threshold and increase the fraction of the population that do not engage in protest. When the public's prior beliefs are bad (i.e. when $\mu_x > \bar{x}$, namely when prior beliefs about level of protest is greater than threshold), the government engages in media manipulation, through either distortion or diversion. By contrast, when the public's prior beliefs are good (i.e. when $\mu_x < \bar{x}$), the government does not necessarily conduct media manipulation. In particular, diversion actively increases the protest fraction, by flattening the distribution and increasing the CDF. The government may choose to engage in distortion contingent in certain cases contingent upon precise values of μ_x, σ_x^2 , as discussed in above section.

For sake of illustration, temporarily disregard the consequences of manipulation on public trust; when the government focuses on lowering probability of unrest, the government's objective is to minimize $1 - \text{CDF}$ of the posterior belief distribution up to the cutoff threshold $P(y > \bar{x})$. So the

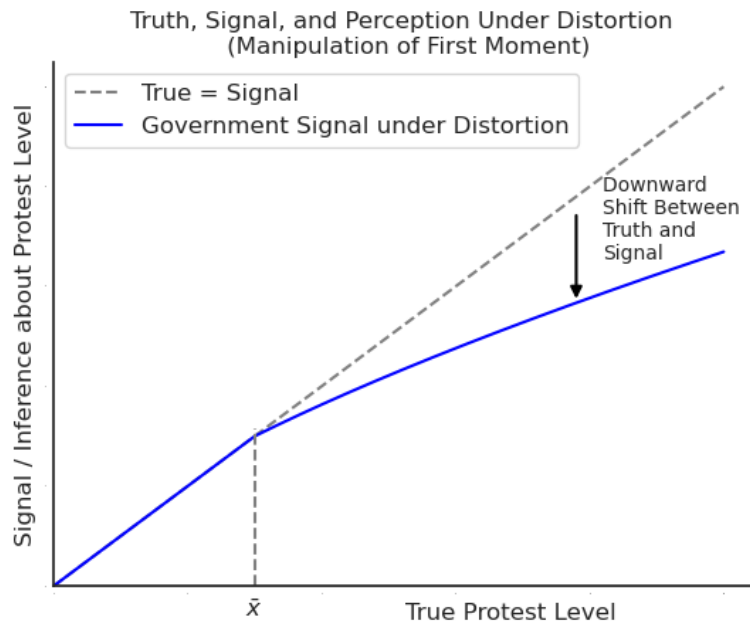


Figure 5: True vs. Perceived Signal Under Distortion (Manipulation of First Moment)

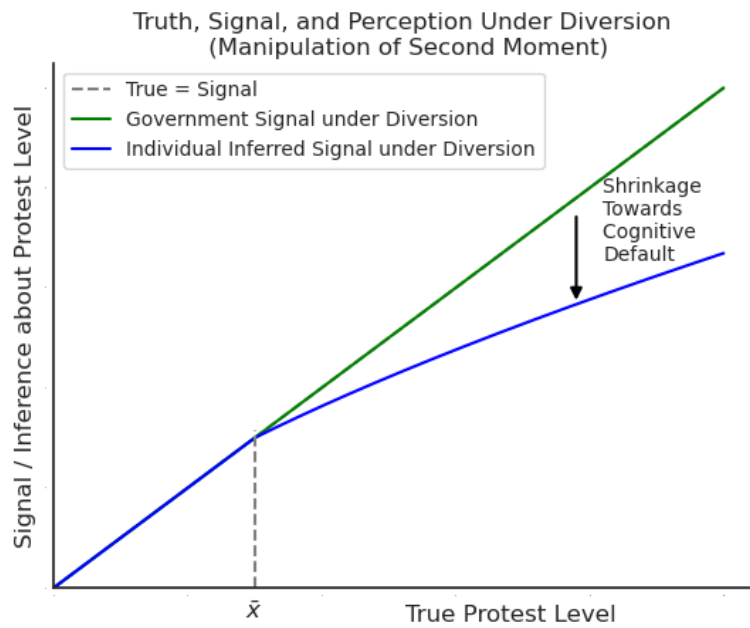


Figure 6: True vs. Perceived Signal Under Diversion (Manipulation of Second Moment)

government solves

$$\min 1 - F_y(\bar{x}) = P(y \geq \bar{x}) = 1 - \Phi\left(\frac{\bar{x} - \mu_y}{\sigma_y}\right) = \int_{\bar{x}}^{\infty} y f_Y(y) dy$$

Figure 7 shows that during bad states of the world ($\mu_x > \bar{x}$), the government can decrease $1 - \text{CDF}$ (the shaded region) by either flattening the posterior distribution or shifting the posterior distribution downward. By contrast, as shown in Figure 8 during good states of the world ($\mu_x < \bar{x}$), diversion actually counter-productively increases the shaded area, reaffirming the earlier intuition that the government only adds noise for the purpose of shading the posterior upward.

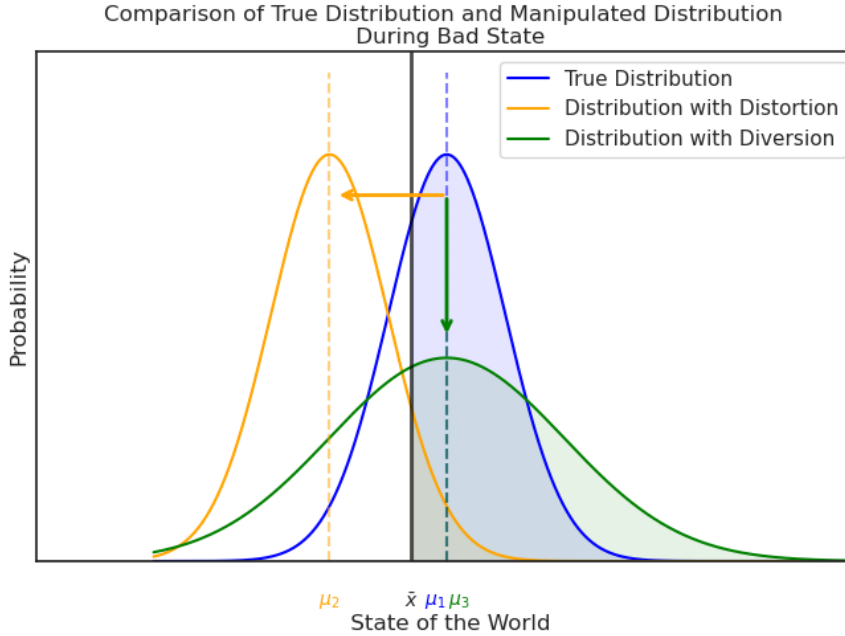


Figure 7: True vs. Manipulated Distribution in Bad state of the World

2.7 Operationalizing Distortion and Diversion through Weibo: A Toy Example

Next I explicitly model how distortion and diversion operate mechanistically through the news media channel on Weibo. A state-run media posts N total posts within a given timeframe $\{1, 2, \dots, N\}$. Note N is relatively fixed and mostly does not fluctuate drastically across time; however, N can vary from day to day within a small range. Among the N articles, news vary in characteristics along three dimensions: 1) location (domestic vs. foreign), 2) sentiment (positive vs. negative), and 3) topic (political vs. nonpolitical). There are N_p political news and N_e

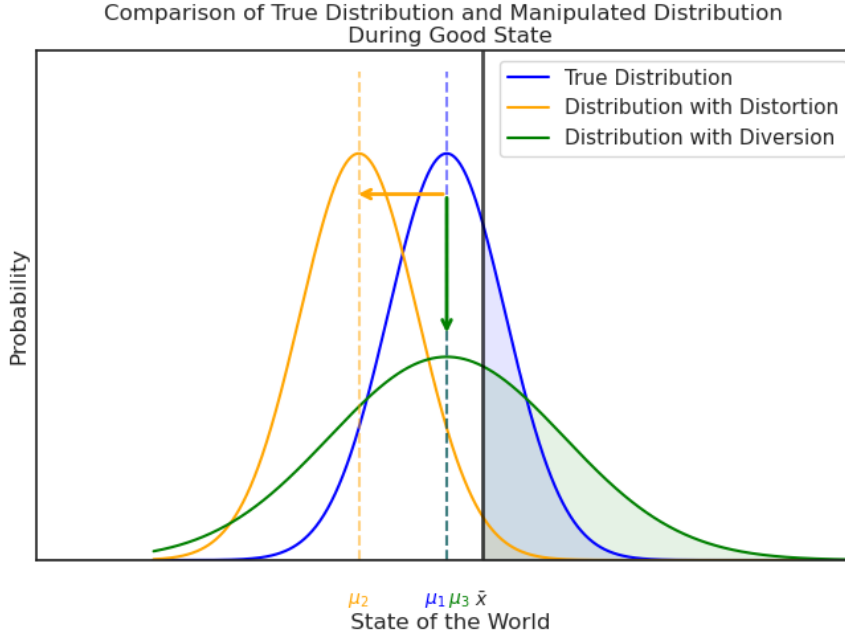


Figure 8: True vs. Manipulated Distribution in Good state of the World

entertainment (i.e. non-political) news, N_g positive (good) news and N_b negative (bad) news, and N_d domestic news and N_f foreign news, such that each binary category partitions the sample space.

$$\begin{cases} N_p + N_e = N \\ N_g + N_b = N \\ N_d + N_f = N \end{cases}$$

Individuals consume media in a way similar to drawing balls from an urn: the probability of drawing a news of type i is proportional to the number of news in the category i and inversely proportional to the number of news in other categories.

There are various ways for the government to manipulate state-run media to generate distortion and diversion (i.e. change μ_d and σ_d). Tables 1 and 3 explicitly model how censorship, propaganda, embellishment, fabrication, and diversion work in the context of news posts. Tables 1 and 2 present a concrete toy example of manipulated media and free media. Tables 3 and 4 summarize the directional shifts under each circumstance.

I distinguish between a baseline (counterfactual) *without* government manipulation of media content (which resembles outcomes for a free and independent media) and the scenario of interest

When the government ENGAGES in media manipulation						
	N_p				N_e	Total N
	$N_{p,d}$		$N_{p,f}$			
	$N_{p,d,b}$	$N_{p,d,g}$	$N_{p,f,b}$	$N_{p,f,g}$		
Baseline (Average Day With No Unrest)	a	b	c	d	e	N
Artificial Suppression of Reportage of Negative Domestic Political events leaving all else constant	$a - \Delta$	b	c	d	e	$N - \Delta$
Artificial Sentiment Shift from Negative to Positive	$a - \Delta$	$b + \Delta$	c	d	e	N
Artificial Location Shift from Domestic to Foreign	$a - \Delta$	b	$c + \Delta$	d	e	N
Artificial Topic Shift from Political to Nonpolitical (while changing the ratio of positive vs. negative domestic political news)	$a - \Delta$	b	c	d	$e + \Delta$	N
Artificial Topic Shift from Political to Nonpolitical (while holding the ratio of positive vs. negative domestic political news)	$a - \Delta$	$b - \Delta$	c	d	$e + 2\Delta$	N

Table 1: Distribution of Number of News Reporting Under Government Media Manipulation

with government manipulation of media. Comparing the predicted direction of change (predicted sign of regression coefficient for each of the manipulation channels), we see that the signs are perfectly symmetric/flipped for cases *with* and *without* manipulation. Table 3 summarizes the observable outcomes of media posts when government manipulation is present, while table 4 describes when government manipulation is absent. For cases where manipulation is present, I additionally distinguish between different conceptual mechanisms of government manipulation, and model how each channel would be operationalized. For cases where manipulation is absent, I also enumerate different mechanisms through which the composition of posts would change.

When a negative domestic political event occurs, the total number of Weibo posts N may stay the same or slightly increase. I describe both channels in the table; while they differ in terms of the absolute levels, they are identical in relative proportional terms (which are our regression outcome variables of interest).

When the government DOES NOT ENGAGE in media manipulation						
	N_p				N_e	Total N
	$N_{p,d}$		$N_{p,f}$			
	$N_{p,d,b}$	$N_{p,d,g}$	$N_{p,f,b}$	$N_{p,f,g}$		
Baseline (Average Day With No Unrest)	a	b	c	d	e	N
Natural Increase in Reportage of Negative Domestic Political events leaving all else constant	$a + \Delta$	b	c	d	e	$N + \Delta$
Natural Sentiment Shift from Positive to Negative	$a + \Delta$	$b - \Delta$	c	d	e	N
Natural Location Shift from Foreign to Domestic	$a + \Delta$	b	$c - \Delta$	d	e	N
Natural Topic Shift from Nonpolitical to Political	$a + \Delta$	b	c	d	$e - \Delta$	N

Table 2: Distribution of Number of News Reportings Under Free and Independent Media

When the government ENGAGES in media manipulation							
	Effect on Posterior Distribution Y		Effect on Outcome Variables (Sign of Regression Coefficient)				
	μ_y	σ_y^2	$\frac{N_{p,d,b}}{N_{p,d}} \propto 1 - \mu_y$ (Fraction Negative Political Domestic Given Political Domestic)	$\frac{N_{p,b}}{N_p} \propto \frac{1}{\sigma_y^2}$ (Fraction Domestic Given Political)	$\frac{N_{p,b}}{N}$ (Fraction Negative Political)	$\frac{N_{p,d,b}}{N}$ (Fraction Negative Political Domestic)	$\frac{N_p}{N}$ (Fraction Political)
Baseline (Average Day With No Unrest)	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Artificial Suppression of Reportage of Negative Domestic Political events leaving all else constant	Increase	Increase	Decrease	Decrease	Decrease	Decrease	Decrease
Artificial Sentiment Shift from Negative to Positive	Increase	No Change	Decrease	No Change	Decrease	Decrease	No Change
Artificial Location Shift from Domestic to Foreign	Increase	Increase	Decrease	Decrease	No Change	Decrease	No Change
Artificial Topic Shift from Political to Nonpolitical (while changing the ratio of positive vs. negative domestic political news)	Increase	Increase	Decrease	Decrease	Decrease	Decrease	Decrease
Artificial Topic Shift from Political to Nonpolitical (while holding the ratio of positive vs. negative domestic political news)	No Change	Increase	No Change	Decrease	Decrease	Decrease	Decrease

Table 3: Conceptual Effects on Posterior Distribution and Outcome Variables (In Presence of Media Manipulation)

When the government DOES NOT ENGAGE in media manipulation							
	Effect on Posterior Distribution Y		Effect on Outcome Variables (Sign of Regression Coefficient)				
	μ_y	σ_y^2	$\frac{N_{p,d,b}}{N_{p,d}} \propto 1 - \mu_y$ (Fraction Negative Political Domestic Given Political Domestic)	$\frac{N_{p,b}}{N_p} \propto \frac{1}{\sigma_y^2}$ (Fraction Political Domestic)	$\frac{N_{p,b}}{N}$ (Fraction Negative Political)	$\frac{N_{p,d,b}}{N}$ (Fraction Negative Political Domestic)	$\frac{N_p}{N}$ (Fraction Political)
Baseline (Average Day With No Unrest)	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Natural Increase in Reportage of Negative Domestic Political events leaving all else constant	Decrease	Decrease	Increase	Increase	Increase	Increase	Increase
Natural Sentiment Shift from Positive to Negative	Decrease	No Change	Increase	No Change	Increase	Increase	No Change
Natural Location Shift from Foreign to Domestic	Decrease	Decrease	Increase	Increase	No Change	Increase	No Change
Natural Topic Shift from Nonpolitical to Political	Decrease	Decrease	Increase	Increase	Increase	Increase	Increase

Table 4: Conceptual Effects on Posterior Distribution and Outcome Variables (In Absence of Media Manipulation)

As shown in the table, our major outcome variables of interest are $\frac{N_{p,d,b}}{N_{p,d}} \propto 1 - \mu_y$ (Fraction Negative Political Domestic Given Political Domestic), $\frac{N_{p,d}}{N} \propto \frac{1}{\sigma_y^2}$ (Fraction Political Domestic), $\frac{N_{p,d,b}}{N}$ (Fraction Negative Political Domestic), and $\frac{N_p}{N}$ (Fraction Political).

I note the mechanical linkage between the proportions observable from the empirics and the first and second moments of the posterior belief distribution derived from the model. The signal in question concerns “how good or bad is the state of affairs for domestic political issues.” The first moment μ_y in the model represents the average level of sentiment (positiveness), which is equivalent to $1 - \frac{N_{p,d,b}}{N_{p,d}}$ (the fraction of political domestic posts that are positive) in the data. The second moment σ_y^2 represents the variance of the posterior, which is inversely proportional to $\frac{N_{p,d}}{N}$ (the fraction of all posts that are about the relevant topic, i.e. domestic politics) in the data.

Conveniently, table 2 supplies a benchmark for comparison, as it describes the counterfactual scenario of the media landscape absent government intervention; namely, a free and independent media theoretically behaves this way. Noticeably, the predicted regression coefficients have signs opposite from those in the government-intervened world.

Since the predicted signs of change across the four specifications are different for all five channels, the manipulation channel used by the Chinese government can be specifically identified from the empirical results, by observing which set of predictions (amongst all five channels) aligns closest with our regression outcomes. In practice, however, the channels likely work in concert, as the government simultaneously shifts attention and conversation towards topics unrelated to bad domestic politics, regardless of whether it is about good domestic politics, foreign politics, or nonpolitical topics. Likewise for the free media benchmark, no one channel is likely to exist in isolation from others.

In Table 3, compared to the regression baseline (which is an average day presumably with no or minimal unrest events), following onset of bad domestic political event:

1. Artificial suppression of reporting of negative domestic political events (leaving all else constant) increases both the mean μ_y and variance σ_y^2 of the posterior belief; specifically, it increases the mean by decreasing the fraction political domestic posts that are also negative (i.e. removing the negative signals from the pool of relevant signals, which lifts up the average value of the signal), and increases the variance by decreasing the fraction of total posts that are political

and domestic (i.e. decreasing the number of signals, which statistically increases the variance of the signal). As a byproduct, it also decreases the overall fraction of political posts, as well as the fraction of negative political posts.

2. Artificial sentiment shift from negative to positive news increases the mean μ_y but does not affect the variance σ_y^2 of the posterior belief; specifically, it increases the mean by decreasing the fraction political domestic posts that are also negative, and leaves the variance unchanged by not affecting the fraction of total posts that are political and domestic (i.e. since it simply replaces negative signals with positive signals, it does not change the total number of domestic political signals and hence the variance). Likewise, it also exerts no effect on the overall fraction of political posts.
3. Artificial location shift from domestic to foreign increases both mean μ_y and variance σ_y^2 of posterior beliefs; it increases the mean by decreasing the fraction political domestic posts that are also negative (i.e. removing negative news on domestic politics and leaving the number of positive news on domestic politics unchanged), and increases the variance by decreasing number of domestic political news.
4. Artificial topic shift from political to nonpolitical (while changing the ratio of positive vs. negative domestic political news) increases both mean (by reducing number of bad domestic political signals) and variance (by reducing the number of domestic political signals overall).
5. Artificial topic shift from political to nonpolitical (while holding the ratio of positive vs. negative domestic political news constant by simultaneously lowering the positive domestic political news) does not change the mean (as the ratio of good to bad news about domestic politics is kept constant by removing both good and bad news in equal proportions) but increases the variance (as number of signals about domestic politics are doubly reduced).

This is compared to the free media benchmark where reportage on negative domestic political events would naturally increase following the occurrence of a negative event in domestic politics, which decreases μ_u and σ_y^2 in most cases. The comparison between free and manipulated media squares with intuition. Following the onset of a bad domestic political event, the signals sent by the free media shifts downward expectations about the state of the world and reduces variance in the

posterior (since individuals now possess more information). To the contrary, a government managed media sends signals that either shifts expectations upward or at minimum leave it unchanged, and inflates the variance (since reduction in reporting leaves individuals with less information to arrive at the correct inference about the prior distribution).

3 Data Collection and Processing

3.1 Measuring Political Unrest: GDELT Data

I construct a time-series index of social unrest activity, sourcing the data from the Global Database of Events, Language, and Tone (GDELT) Project. Developed by Google, GDELT monitors the world’s news media around the world in broadcast and web formats.¹ GDELT uses data mining and natural language processing algorithms to extract events and their corresponding themes and emotions. GDELT records instances of events based on articles from a global, comprehensive set of media outlets.² Each event identifies a source actor, an event type, and a target actor. When multiple news sources cover the same event, GDELT records only one event. The GDELT dataset is updated in real time with 15 minute intervals, and contains over one hundred news sources that provide coverage on China.³ While the Chinese data coverage may be biased by government censorship, it offers one of the most comprehensive information on China.⁴

The GDELT project covers over 300 categories related to political activity; each event falls under the following general umbrellas.⁵ The various event categories include: make public statement, appeal, express intent to cooperate, consult, engage in diplomatic cooperation, engage in material cooperation, provide aid, yield, investigate, demand, disapprove, reject, threaten, protest, exhibit force posture, reduce relations, coerce, assault, fight, and use unconventional mass violence.

The dataset includes the universe of events between 2014 and 2020 where both the actor and target are geocoded to be within China. For each event, the following variables are pertinent to our analysis:

¹The GDELT data is graciously provided by Prof. David Yang and Andrew Kao.

²GDELT creates two primary data streams: one codifying physical activities around the world into over 300 categories, and another recording the people, places, organizations, themes, and emotions underlying those events.

³For detailed information on GDELT, refer to their website.

⁴See Beraja et al. (2021) for a more detailed discussion.

⁵GDELT adopts the CAMEO version 1.1b3 taxonomy. For more detailed information, please refer to the CAMEO User Manual or the GDELT website for crosswalk files

- **Quad Class:** Specifies this primary classification for the event type, allowing analysis at the highest level of aggregation. The four primary classifications include: Verbal Cooperation, Material Cooperation, Verbal Conflict, and Material Conflict.
- **Goldstein Score:** A numeric score from -10 to +10 capturing the theoretical potential impact that type of event will have on the stability of a country, using the Goldstein Scale. (Note: the score does not account for size of protest)
- **Number of Article Mentions:** the total number of source documents containing one or more mentions of this event during the 15 minute update in which it was first seen. This can be used as a method of assessing the “importance” of an event: the more discussion of that event, the more likely it is to be significant.
- **Average Tone:** This is the average “tone” of all documents containing one or more mentions of this event during the 15 minute update in which it was first seen. The score ranges from -100 (extremely negative) to +100 (extremely positive). Common values range between -10 and +10, with 0 indicating neutral.
- **Location:** Since GDELT data is geographically referenced to the FIPS10 administrative division 1 (ADM1) level (provincial level), this allows us to observe the specific location at which these events occurred, and connect unrest data with weather data.

We additionally construct the following variables.⁶

$$\text{Coverage Weighed Event Tone} = \text{Average Tone} \times \log(1 + \text{Number of Article Mentions}) \quad (15)$$

$$\text{Destabilization Potential} = 0.8 \times \text{Goldstein Score} + \text{Coverage Weighted Event Tone} \quad (16)$$

Summary statistics and figures are displayed below. Most protests in China are small-scale, localized, and short-lived, and focus on economic issues. Concerning actors, the majority of GDELT events lack either an identified actor or identified target. For the subset of identified events, Figures 25, 26, and 27 exhibit a frequency matrix for the top 7 most frequent actor and target types, which

⁶Note that more negative scores indicate higher destabilization potential

include government, business, civilian, police forces, military, media, and education institutions. Among all the events, nongovernmental entities account for the majority of actors and targets in the events. Nongovernmental entities are the actors in 32.66% of events and are the target in 33.90% of events. When jointly considering the actor and target, cases where nongovernment entities are both the actor and target lead in terms of frequency by a slim margin, accounting for 36.06% of events; cases of government acting upon entities and nongovernment entities acting upon government are similarly frequent at 30%. Meanwhile, events where the government is both the actor and target entity are the least probable at only 2.63%.

Concerning types of event, figure 22 reports the relative frequencies of verbal cooperation, material cooperation, verbal conflict, and material conflict. Of the events, 45.5% are verbal conflicts, 39.9% are material conflicts, and 14.6% are material cooperations. Table 6 displays the relative frequency of events for non government actors, separated by action type and topic type. Of the 31428 events by nongovernmental actors, we observe that the most prevalent types of action are disapproval and fight. Demands, threats, and protests are also common.

Figure 14 and Table 21 present the geographical distributions of unrest events: Beijing and Gansu feature the highest event counts; however, their respective profiles of and causes for unrest are starkly different. For Gansu, verbal and material conflicts are more common as a fraction of all events, whereas the contrary is true for Beijing: Gansu’s 50.2% frequency of verbal conflicts is higher than the national average of 45.5%. For Beijing, the government represents a higher than national average fraction of events (37% compared to 32%), while for Gansu, civilians and private entities initiate a higher proportion of events (70.65% compared to 67.4%). The case of Gansu illuminates various factors that contribute to social unrest. Economically, Gansu is the nation’s poorest province with the lowest GDP per capita; low agricultural productivity and high unemployment rate contribute to economic instability, spurring riots and protests against economic policies such as redistribution of tourism revenue. Societally, Gansu is home to a number of ethnic minority groups, including Tibetans, Mongolians, and Hui Muslims. Ethnically rooted tensions between minorities and the Han majority have also exacerbated unrest. Environmentally, frequent earthquakes, droughts, and famines lead to disputes over land and water resources. Lastly politically, Gansu has a history of corruption, poor governance, and oppressive government response to protests, increasing animosity towards the government. For instance, in November 2018, residents

of a Tibetan township in Gansu gathered to protest a government plan to divert fresh drinking water from a local source to a neighboring area, attracting police forces who threatened them with arrest.

Concerning timing, figures 15 and 16 reveal temporal patterns on a weekly and annual basis. On a micro level, unrest events are more likely to occur during weekdays compared to weekends. This may be attributed to either biased sampling due to greater media coverage on business days than weekends, or to weekdays offering attractive features like higher visibility and greater impact, particularly if protestors target government institutions, political figures, or businesses and aim to disrupt their operations. On a macro level, unrest is less common in October, since the government is particularly focused on maintaining political stability and social harmony around National Day (October 1) via increased security measures and crackdowns on dissent. Unrest is most common in March, April, and July, likely due to the following reasons. First, anniversaries of political movements like the Tibetan Uprising Day (March 10) and the Urumqi riots (July 5) bring attention to issues of human rights, political freedoms, and ethnic tensions. Second, the annual meeting of China’s top legislative body the National People’s Congress in March, as well as the CCP annual leadership meeting in July, draws popular attention to possibly contentious political subjects. Third, nonpolitical forces like seasonal employment starting in March after the Lunar Festival can spark labor market frictions particularly for seasonal migrant workers. Lastly, milder weather conditions in spring and summer months are more conducive to outdoor assemblies.

Regarding the impact potential of events, figures 17 to 21 plot the event level distribution of goldstein scale, average tone across reportings, and number of article reportings, as well as constructed measures coverage weighted event tone and destabilization potential, all of which proxy for the potential societal and political impact. Destabilization potential is decreasing in both tone (measure of affect) and goldstein score (measure of intensity) and increasing in the number of article mentions (measure of magnitude of public attention).

Specifically, figures 28 and 29 display the goldstein score and average tone for events by actor and target type. In accordance with intuition, events of type “civilian unto military,” “police unto civilian,” “police unto police,” and “civilian unto police” have the lowest average tone across media reporting. Similarly, events of type “civilian unto business” , “military onto media” are the most destabilizing as suggested by the Goldstein score.

3.2 Measuring Government Media Activity: Weibo Data

The Chinese government can modify people’s perception of the state of the world by altering the information they are exposed to; social media platforms like Weibo furnish a useful empirical context. I construct a novel dataset that includes the Weibo social media accounts of China’s top 18 national-level state-run media. I automate a web scraping algorithm to search the corresponding social media accounts and scrape their account activity. Given this dataset, I next utilize a supervised machine learning topic classification model to categorize the posts’ content along various dimensions including location, affect, attention-grabbingness, and topic label to produce a feature dataset.

I examine state-run media posting activity on Weibo as an empirical setting, for a few reasons. First, since Chinese state run media are government official media outlets, they allow us to observe the direct action of the government, absent concerns about private sector media activity. Studying state run media is preferable to studying overall aggregate Weibo activity, as the latter only allows us to observe an equilibrium outcome involving both private actors and the government (namely, both individual self censorship and government censorship/propaganda/information regulation are involved), whereas the former isolates the government’s media activity, thereby obviating concerns about decomposing equilibrium outcomes. Second, social media operates on a high frequency and allows for instantaneous posting and viewership. As such, it’s an appealing mechanism for the government to divert attention, as content can be instantaneously and costlessly (in pecuniary terms) inserted, removed, or altered. Thus, compared to slower-moving media industries like TV shows and films which require a long time to orchestrate and produce, Weibo appears as a plausible mechanism for the government to respond to political unrest activity due to the short time horizon. Third, Weibo is a heavily utilized social media platform in China, akin to Twitter in the United States. Due to its wide audience and ubiquitous presence, Weibo presents itself as a cost-efficient and influential medium for the government to interface with the public.

I scrape a total of 1.2 million Weibo posts from 18 Chinese media outlets spanning from 2011 to 2023, including *People’s Daily*, *Xinhua News*, *Economic Daily*, *China Daily*, etc. The state-run media are summarized in Table 11. Figure 13 plots the number of posts per day across time. I observe that generally the posting activity is stable across time for all media outlets, excepting

upticks during extraordinary periods like the onset of COVID.

For each post, I extract the following variables: text, date and time, presence of audio-visual media (i.e. whether the post contains images and videos), number of likes, number of comments, number of reposts, and whether the post is original or reposted. In particular, I focus on three aspects of the data: 1) the Weibo text for topic classification purposes, 2) the presence of audio-visual media as a candidate measure for attention-grabbingness, and 3) like, comment, and repost activity as audience engagement and attention towards the post.

3.2.1 Topic Classification

BERT I first train a Bidirectional Encoder Representations from Transformers (BERT) model on a pre-existing dataset and next fine-tune it on a hand labeled dataset. In broad terms, the model performs feature selection, feature dimension reduction, and classification model learning.

I overcome a few challenges with the dataset. A first concern is that the dataset is sparse and binary topic labels are unbalanced (i.e. while common topics like politics and economy feature large numbers of posts, less common topics like sports and environment feature many 0's and few 1's). I address this in the following two ways: First, in terms of training the classifier model, I reweigh the training set to overweight 0 labels relative to 1 labels (these weights are multiplied with the positive rate to make the dataset balanced between number of 0's and number of 1's). I additionally adjust the decision boundary/threshold of the sigmoid activation function from 0 to a positive value, which induces the model to classify texts as 0 with greater probability. Second, in terms of evaluating model performance, I examine false positive and false negative rates in addition to accuracy, as accuracy is a suboptimal measure of model performance in the case of sparsity and unbalanced labels.

Transfer learning Since our hand-labeled dataset is limited in size, I train the entire network on the pre-labeled THUNews dataset, and finally fine tune the last layer of our neural network on our hand-labeled dataset. THUCNews is a dataset of historical filtered news texts from Sina News RSS Subscription Channel for the period 2005-2011, provided by the Natural Language Processing Laboratory of Tsinghua University. The dataset includes 740,000 news articles, with a size of 2.19GB. On the basis of the original systematic news categorization by Sina News, the data is

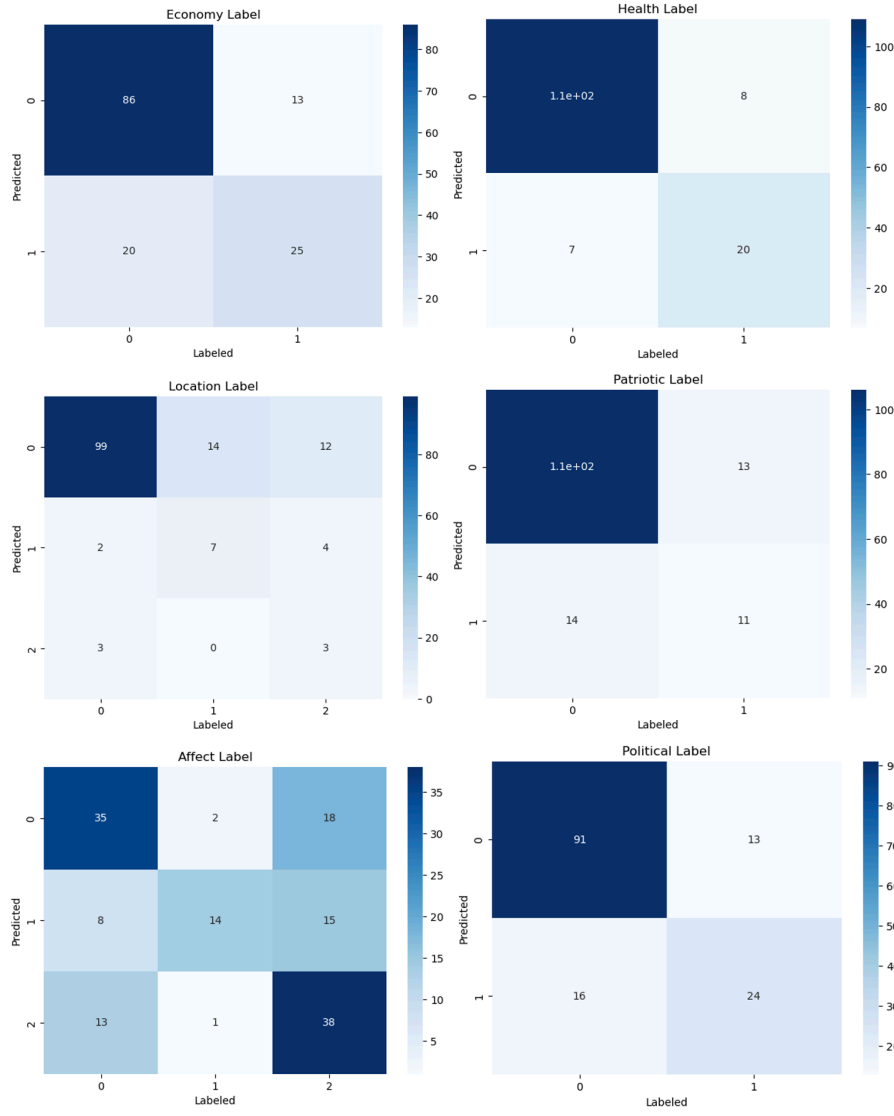


Figure 9: Confusion Matrix for Topic Classification Labels

In each confusion matrix, 0 represents a negative label, 1 represents a positive label; top left cell represents true negatives, top right represents false negatives, bottom left represents false positives, and bottom right represents true positives.

Topic Label	Accuracy
Location	77.0
Affect	59.8
Political	80.5
Patriotic	81.6
Attention Grabbingness	63.2
Politics and Government	75.9
Economy	85.1
Law, Crime, and Justice	87.4
Disaster and Accident	92.0
Health	82.8
Culture	90.8
Society	75.9
Entertainment, Media, and Arts	87.4
Lifestyle and Leisure	73.6
Environment	83.9
Education	93.1
Science and Technology	88.5
Sport	93.1

Table 5: BERT Model Topic Classification Accuracy Statistics by Topic Label

further categorized into 14 candidate categories, including finance, lottery, real estate, stocks, home, education, technology, society, fashion, politics, sports, horoscope, gaming, and entertainment. THUCNews provides us with a wealth of information to train our model.

Weibo Training Dataset Our Weibo dataset includes 1500 randomly selected Weibo texts, and is therefore a representative sample. I use a 90-10 train-test split, resulting in 1350 posts for training and 150 posts for testing. I manually classify each post along the following dimensions

1. Location (Multiclass: Domestic, Foreign, Bilateral/Global, No Region)
2. Tone (Multiclass: Positive, Negative, Neutral)
3. Political (Binary: Political or non-political)
4. Patriotic (Binary: Patriotic or non-patriotic)
5. Attention Grabbing (Binary: Attention grabbing or not; note attention grabbing can include funny, shocking, sensational, etc.)
6. Multibinary Topic Classification between 13 topics

I classify topics using a system adapted from the International Press Telecommunications Council (IPTC) Media Topics NewsCodes, shown in figure 10. Figure 7 shows the summary statistics of the categorization output. Figure 9 presents the confusion matrix (Type-I and Type-II errors) of the predictions relative to manual categorization. Table 5 summarizes the accuracy for all labels. I extract and concatenate features from both the THUNews and Weibo datasets, utilizing both for prediction.

3.3 Instrumenting for Unrest: Weather Data

I use local weather conditions to instrument for political unrest. Historical weather data is collected by the World Meteorological Organization (WMO) and hosted by the National Oceanic and Atmospheric Administration (NOAA). Data is reported at the weather station-day level and includes an array of features like mean temperature, amount of precipitation, presence of fog or hail or thunder, maximum wind speed recorded, and visibility. We then match weather stations with prefectures.

3.4 Variable Construction and Method of Aggregation

Media Data Variable Construction Each Weibo post is classified along the following dimensions: 1) location: domestic vs. foreign vs. global/no location, 2) affect: positive vs. negative vs. neutral, 3) politicalness: political vs. nonpolitical, 4) patriotic vs. non patriotic, 5) multibinary topic classification. This allows us to construct additional binary labels from intersections between orthogonal labels. For instance, we construct “negative domestic political”

$$\textit{Negative domestic political} = \textit{Negative} \times \textit{Domestic} \times \textit{Political} \quad (17)$$

We similarly define other variables like “neutral (positive) domestic political,” “foreign political,” etc.

We also construct the variable log audience attention, which measures how much the audience engages with the post online; compared to comments, we allocate less weight to likes and more weight to shares, as retweeting a post indicates greater attention and engagement with the content than likes or comments. Furthermore, shares are less prone to bot activity compared to likes and comments, as King, Pan and Roberts (2017) documents how the Chinese government historically

employs commentators to flood the comment section with positivity. Audience attention is defined as follows:

$$\log(\text{Audience Attention}) = \log(\text{likes}^{0.8} + \text{comments} + \text{shares}^{1.2} + 1) \quad (18)$$

Unrest Data Variable Construction Each unrest event is classified, amongst other things, by actor type, target type, topic type, and action type. In terms of actor entities and target entities, we collapse specific types crudely into government vs non government entities. Topic types include political, economic, and military. Action types include ease, demand, disapprove, threat, protest, posture, coerce, and fight. This allows us to consider a plethora of different configurations between actor, action, topic, and target. For instance, we consider the permutation and construct “Nongov Threat Gov” as follows.

$$\text{Nongov Threat Gov} = \begin{cases} 1 & \text{if } \text{Actor} = \text{Nongov}, \text{Action} = \text{Threat}, \text{and } \text{Target} = \text{Gov} \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

$$\text{Nongov Demand Political} = \begin{cases} 1 & \text{if } \text{Actor} = \text{Nongov}, \text{Action} = \text{Demand}, \text{and } \text{Topic} = \text{Political} \\ 0 & \text{otherwise} \end{cases} \quad (20)$$

We similarly define other variables like “Nongov coerce nongov,” etc. Furthermore, we create categorical measures of event hostility.

$$\begin{cases} \text{Conciliatory} = \{\text{Ease}\} \\ \text{Nonviolent and Hostile} = \{\text{Demand}, \text{Disapprove}, \text{Threat}, \text{Protest}\} \\ \text{Violent and Hostile} = \{\text{Posture}, \text{Coerce}, \text{Fight}\} \end{cases}$$

Aggregation The richness and granularity of the media and unrest datasets confer the opportunity for flexible aggregation along various dimensions to comprehensively summarize the data. For both the Weibo and the unrest data, we sum counts for all categories to the day level to gauge aggregate media activity and unrest activity for a given day. I aggregate data at the day level for two reasons: First, in terms of data availability, unrest and weather data exists in daily frequency and Weibo

data is precise to the minute, therefore, daily aggregation preserves the unique granularity that would be otherwise lost in weekly aggregation. Second, in terms of the conceptual transmission mechanism, the government should be acutely aware of unrest events and respond swiftly through shifting the composition of catered content, corroborating our modeling decision.

For Weibo and the unrest data, we respectively sum counts for posts and events across all categories to the day level to measure aggregate media activity and unrest activity for a given day; for instance, we tally counts of unrest for each interacted category of actor entity type, target entity type, action type, and topic type for GDELT. However, since not all Weibo posts nor all unrest events are created equal, we consider measures of intensity in addition to count. For Weibo, we calculate sum log audience attention across all posts, as well as for different subsets by topic. For GDELT, we calculate average destabilization potential, goldstein scale, and tone, as well as weighted averages weighted by number of article mentions. We also differentiate between mild and severe events in the following three ways. First, we count events for the subset of goldstein score less than -4, negative destabilization potential, and negative coverage weighted average tone. Second, we calculate a weighted count of events adjusting for destabilization potential. Third, we aggregate more granular indicator variables for action types of varying severity, differentiating between moderate events like demand and disapproval from more militant events like assault and coercion. Fourth, we create counts of unrest events across different quartiles of destabilization potential.

4 Empirical Methodology

We conduct two main categories of empirical analyses to answer two questions. First, do Chinese state-run media respond to unrest by engaging in media manipulation and information management? Second, does media manipulation through distortion and diversion effectively direct public attention away from unrest? Namely, we explore first whether state-run media manipulates its content, and next whether the interventions are effective at achieving its goal of attention diversion.

Nongov Actor (Event Count)		Topic of Action				
Type of Action		Political	Econ	Force	Other	All
	Ease	Ease Political (26)	Ease Economic (114)	Ease Force (1058)	Ease Other	Ease (1198)
	Demand	Demand Political (475)	Demand Economic (15)	Demand Force (19)	Demand Other (3532)	Demand (4022)
	Disapprove	Disapprove Political	Disapprove Economic	Disapprove Force	Disapprove Other	Disapprove (22301)
	Threat	Threat Political (125)	Threat Economic (75)	Threat Force (1094)	Threat Other (4727)	Threat (4927)
	Protest	Protest Political (42)	Protest Economic (313)	Protest Force (125)	Protest Other (2810)	Protest (3290)
	Posture	Posture Political	Posture Economic	Posture Force (1553)	Posture Other	Posture (1553)
	Coerce	Coerce Political (621)	Coerce Economic (608)	Coerce Force	Coerce Other	Coerce (1229)
	Assault	Assault Political	Assault Economic	Assault Force (3280)	Assault Other	Assault (3280)
	Fight	Fight Political	Fight Economic	Fight Force (10816)	Fight Other	Fight (10816)
	All	Political (1289)	Economic (1125)	Force (17945)	Other (11069)	All (31428)

Table 6: Counts of Nongovernment Actor Unrest Events Across Action Type and Topic Type

Topic	Ratio of Negative to Positive Labels
Politics and Government	1.667286
Economy	3.965398
Society	4.199275
Health	5.23913
Lifestyle and Leisure	5.932367
Law, Crime, and Justice	6.396907
Culture	10.299213
Disaster and Accident	10.572581
Environment	13.068627
Education	16.289157
Science and Technology	20.742424

Table 7: Weibo Topic Label Distribution (Manually Classified Sample)

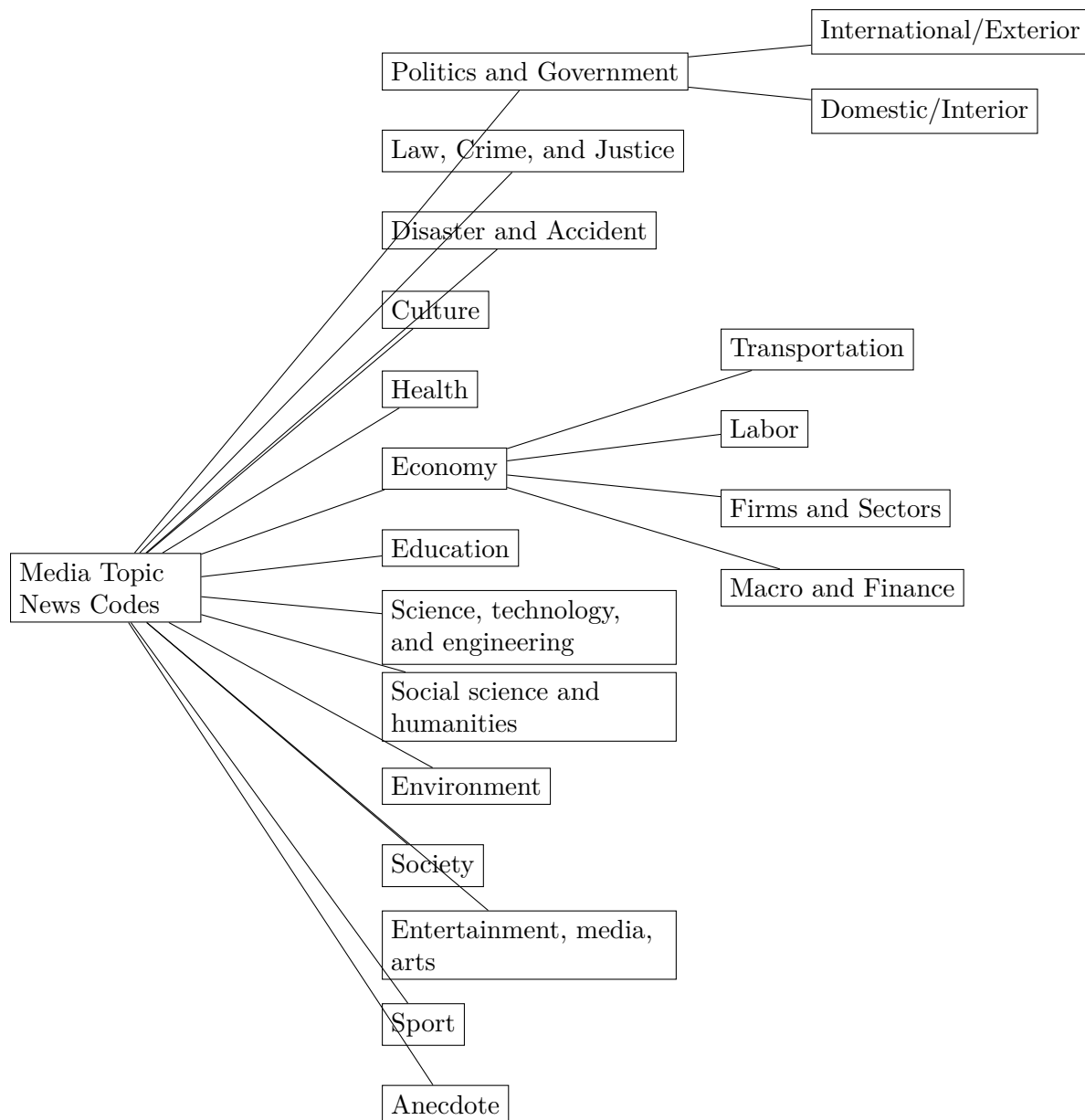


Figure 10: IPTC Media Topic Classification System

4.1 Does the Chinese Government respond to Unrest by Engaging in Diversionary Media?

To investigate whether state-run media change the composition and frequency of media activity following periods of rising unrest, I employ both OLS fixed effect and instrumental variables 2SLS regressions.

4.1.1 OLS Fixed Effects Regressions

We implement preliminary regressions of the following general form, which features media activity as the outcome variable, and exogenous unrest shock as the explanatory variable.

$$\text{Content Share}_{i,t} = \beta_0 + \beta_1 \text{Unrest Level}_{i,t} + \beta_2 \text{Unrest Level}_{i,t-1} + \beta_3 \text{Unrest Level}_{i,t-2} + \delta \mathbf{W}_t + \gamma_i + \epsilon_{i,t} \quad (21)$$

where i indexes the state run media, t indexes time, and γ_i are media fixed effects. *Content Share* _{i,t} is a fractional variable that measures the proportion of posts that are a given topic. *Unrest Level* is the count of unrest events; to explore the heterogeneity of government responses to different event types, we consider various subsets of unrest activity. We include control variables like month of year, day of week, and a linear time trend, to account for the fact that there may be systematic variation in media behavior across times. To control for unobserved heterogeneity that varies across media but remains constant over time, we include media fixed effects, which remove time-invariant differences between entities that may be correlated with the outcome variable. Accordingly, we cluster our standard errors at the media level.

Independent Variables Since not all unrest events are created equal, we examine various dimensions of unrest activity, including actor, target, action, and topic, and estimate separate specifications for each. Due to multicollinearity between the various measures of unrest, we let regressors enter separately instead of jointly. Multicollinearity biases the standard errors upward, and causes unstable patterns in coefficient estimates whose signs may easily flip; therefore, since some fractions are definitionally linear combinations of others, we separately examine the impact of each.

Dependent Variables Our dependent variables are (scaled) fractions that measure the proportion of posts of a certain topic. They fall within the following three umbrellas.

1. Simple Proportion : *Fraction Domestic* ($\frac{\text{Number of Domestic Posts}}{\text{Number of Total Posts}}$), *Fraction Foreign*, *Fraction Positive*, *Fraction Neutral*, *Fraction Negative*, *Fraction Political*, etc.
2. Conditional Proportion: *Fraction Negative Given Domestic Political* ($\frac{\text{Number of Negative Domestic Political Posts}}{\text{Number of Domestic Political Posts}}$), *Fraction Domestic Given Political* ($\frac{\text{Number of Domestic Political Posts}}{\text{Number of Political Posts}}$), etc.
3. Interacted Proportion: *Fraction Negative Domestic Political* ($\frac{\text{Number of Negative Domestic Political Posts}}{\text{Number of Total Post}}$), *Fraction Apolitical and Patriotic* etc.

I choose the proportion/fraction as our outcome variable of interest, for two reasons. First, in accordance with the limited attention model setup, individuals consume media subject to a time constraint, isomorphic to an attention constraint, as practice, netizens spend a relatively fixed amount of time on social media. Therefore, when individuals intake content from the Weibo information deluge, it is arguably the relative, not absolute, prevalence of content that dictates the probability said content will be consumed. Second, empirically, the total number of posts per media per day is relatively fixed and does not fluctuate much across time; for instance, *People's Daily* has roughly the same number of posts everyday (around 30). Therefore, considering the relative proportions is isomorphic to considering absolute levels. Nonetheless, we also consider total counts of posts per day as an outcome variable, as there exists meaningful day to day variation.

Limitations of OLS Fixed Effects OLS estimates may be imprecise due to omitted variable bias and measurement error of the outcome variable; however, neither concern substantially detracts from our results, since omitted variable bias in this case implies the true coefficient estimate should be larger in magnitude, and measurement error in weibo post topic classification does not bias our estimate but rather deflates the p-value.

First, even when controlling for media fixed effects and event-level observables, confounding factors such as other exogenous shocks including breaking news unrelated to unrest and endogenous economic developments challenge internal validity. Eliminating omitted variables by adding controls is difficult, as other covariates like economic conditions are only observed at longer horizons, subjecting our regressions to omitted variables bias. may be subject to omitted variable bias.

Consider the omitted variable of a positive political event. $\alpha_1 = \beta_1 + \beta_2\gamma_2$, where β_1 is the true effect of unrest event on fraction of negative domestic political news, α_1 is the estimated coefficient under omitted variable bias, and γ_2 is the coefficient of positive political progress in the auxiliary regression. in this context, $\alpha_1 < 0$, $\beta_2 < 0$, and $\gamma_2 < 0$, therefore the bias $\beta_2\gamma_2$ is positive, leading us to understate the true effect, and estimating an effect α_1 smaller in magnitude in the negative direction than the true effect β_1 .

Second, measurement error in the dependent variable reduces statistical precision, as the sampling distribution will be more dispersed, hence widening the coefficient's confidence interval. Meanwhile, measurement error in Y does not bias the coefficient point estimates. Furthermore, the statistical significance of our results even in the presence of wider confidence intervals suggests that true statistical significance under precise measurement of Y may be higher.

4.1.2 Instrumental Variables Regression

OLS is only unbiased if conditional mean independence is satisfied, i.e. the independent variable of interest is uncorrelated with the error term; thus, the internal validity may be compromised. Instrumental variable regressions help us obtain a consistent (but still biased) estimator when the conditional mean independence assumption is violated. Since both media and unrest data are high frequency, control variables that match our dependent and independent variables in terms of frequency are difficult to find. While economic indicators like GDP, inflation, and unemployment are informative metrics that could meaningfully explain variation in state-run media activity, their quarterly frequency renders them less effective.

To address omitted variables bias, we exploit exogenous variation in weather conditions in an instrumental variables (2SLS) framework. Weather instruments have been widely employed in the political economy subfield. For example, Madestam et al (2013), Gonzalez (2020), and Larrebourg and Gonzalez (2021) use daily weather shocks during rallies as an exogenous driver of attendance. Due to data limitations and high dimensionality concerns, we restrict the 2SLS exercise to Beijing, leveraging Beijing weather data as exogenous driver for unrest activity and observing content composition of Beijing Daily, the top state-run newspaper in Beijing.

We check instrumental variable assumption of relevance, conditional independence, exclusivity, and monotonicity for the weather instrument. First, we demonstrate the relevance condition, which

states

$$\text{Cov}(\textit{Adverse Weather}, \textit{Unrest}) \neq 0 \quad (22)$$

Economically, this requires the instrument (weather) to be a meaningful determinant of the treatment (unrest). Previous literature demonstrates that adverse weather conditions like precipitation, thunder, and gust significantly affect protest turnout. Following Angrist and Pischke (2009), we test relevance by running first stage regressions (with appropriate controls included to make the instrument as good as randomly assigned) and observing the first stage F-statistic in Table 15. We note that while our instrument clears the critical threshold of 10 proposed by Angrist and Pischke (2009), it does not attain the critical value of 23.01 proposed by Montiel-Olea and Pflueger (2013), suggesting that the instrument is sufficient but not incredibly strong. Moving forward, a better methodology for high dimensionality instrument selection would be to perform LASSO regression selection of weather predictors among all weather variables and their interactions.

$$\text{2SLS First Stage: } \textit{Count of Unrest Events}_t = \beta_0 + \beta_1 \textit{Adverse Weather}_t + \gamma \mathbf{W}_t + \epsilon_t \quad (23)$$

Second, we explain our inclusion of controls and discuss the conditional independence assumption, which states

$$\mathbb{E}[\epsilon_t | \textit{Adverse Weather}_t, \mathbf{W}_t] = \mathbb{E}[\epsilon_t | \mathbf{W}_t] \quad (24)$$

Economically, this requires average weather conditions to be as good as randomly assigned across days, once a complete vector of controls is included, i.e. the instrument does not affect the conditional mean of error terms conditional upon controls. This condition is satisfied since weather conditions are completely exogenously determined by nature.

Third, we show that the exclusion restriction, which states that the instrument can only affect the outcome through its effect on the treatment, holds. Since our main outcome variables are the conditional proportions of political content (i.e. the fraction of domestic political content that is negative in affect), they are not directly or indirectly impacted by weather other than through unrest events. As a thought exercise, suppose severe weather conditions on a certain day cause state-run media to report on it, this does not impact the internal composition of political news across sentiment classes, which bolsters the exclusion restriction. While in the case

of overidentifying restrictions, we can perform the J test of overidentifying restrictions to verify the exogeneity condition, in the just identified case where number of instruments equals number of endogenous regressors, we cannot statistically diagnose exogeneity. Moreover, in the most likely scenario of treatment effect heterogeneity, in which each instrument estimates a different local average treatment effect, the J-test of overidentifying restrictions will only inform us whether the instruments estimate the same LATE, but inconclusive as to exogeneity.

Lastly, we consider the monotonicity assumption, which states

$$\frac{\partial Unrest\ Level}{\partial Adverse\ weather} \leq 0, \forall t \quad (25)$$

This is also plausibly satisfied as protest probability and size should both be monotonically decreasing in adverse weather conditions.

Nonetheless, remaining limitations exist for the generalizability of IV2SLS results. First, IV estimates capture the local average treatment effect, which may differ from the average treatment effect over the entire time range. Second, the sparsity of severe weather conditions implies that LATE is estimated over a relatively small sample of days. When considering both the extensive and intensive margins of political unrest, weather conditions may affect protest on the intensive margin but not the extensive margin. As discussed in Beraja et al. (2021), the elasticity between weather and unrest may be zero on some days and nonzero on other days. Over a relatively long period of time, on the one hand, there are many days on which no unrest takes place, not due to unfavorable weather conditions but due to the absence of need for political mobilization; therefore, on these days, there is no elasticity between weather conditions and unrest occurrence. This poses a limitation.

We directly estimate the second stage regression as follows.

$$2SLS\ Reduced\ Form: \log(Content\ Share_t) = \beta_0 + \beta_1 Adverse\ Weather_t + \gamma \mathbf{W}_t + \epsilon_t \quad (26)$$

$$2SLS\ Second\ Stage: \log(Content\ Share_t) = \beta_0 + \beta_1 Count\ of\ \hat{Unrest\ Events}_t + \gamma \mathbf{W}_t + \epsilon_t\ where\ \beta_1 = \frac{\alpha_1}{\pi_1} \quad (27)$$

4.2 Does government diversionary media effectively divert public attention?

After examining the government’s intention to divert, we shift to investigate the potency of media intervention in diverting attention, i.e. market outcomes of diversion. We implement an ordinary least squares and instrumental variables specification to examine how much the audience’s attention towards and engagement with government posts shift in response to media manipulation around unrest events. For the OLS approach, we regress *Log Audience Attention* against the state-run media content composition variables (e.g. *Fraction Negative Domestic Political*, *Fraction Domestic Political*). For the IV approach, we use level of unrest activity (e.g. count of verbal and material conflicts) to instrument for changes in media content composition.

4.2.1 OLS Fixed Effects Regressions

We estimate three regressions of the following general form:

$$\log(\text{Aggregate Attention}_{i,t}) = \beta_0 + \beta_1 \text{Content Fraction}_{i,t} + \delta \mathbb{Z}_t + \gamma_i + \epsilon_{i,t} \text{ where } \mathbf{W}_t \text{ is a vector of controls} \quad (28)$$

The dependent variable is the log-transformed audience engagement with posts of media i on day t , as measured by audience likes, comments, and retweets. The main treatment independent variable *Content Fraction* is the same as our outcome variable in our previous section. Media fixed effects γ_i and observable controls $\mathbb{W}_t$ are included as usual.

4.2.2 Instrumental Variables Regression

The OLS fixed effects regression may be limited in its potency due to omitted variables bias and reverse causality bias. Omitted determinants of audience engagement include media coverage of other non state-run outlets; in particular, if the audience informs themselves first through other channels and subsequently visits state-run media accounts to seek more information, the causal driver of increased engagement with state-run media is then not their own posting behavior but rather that of other media; this confounding variable likely biases our coefficient estimates upward. Another threat to internal validity arises from reverse causality bias: state-run media may be shifting its content composition in response to changes in audience engagement.

The above two challenges again point to IV2SLS as a preferable identification strategy. The number of verbal and material conflicts provides a plausible instrument, as it fulfills both the relevance and exogeneity requirements. First, first stage regressions indicate strength in all five instruments (unrest measures), by virtue of having heteroskedasticity robust F statistics that unambiguously clear the critical value. Appendix tables show how the *Count Verbal Material Conflict* instrument performs for various endogenous variables, and how various subset instruments perform for the endogenous variable *Log Aggregate Attention Potential*. We conclude that individually, each instrument is strong.

$$2SLS \text{ First Stage: } Content \ Fraction_t = \beta_0 + \beta_1 Count \ of \ Unrest \ Events_t + \gamma \mathbf{W}_t + \epsilon_t \quad (29)$$

Second, the exogeneity condition, which requires the instrument to be uncorrelated with residuals ($\mathbb{E}(u_i|Z_{1i}, Z_{2i}, controls) = \mathbb{E}(u_i|controls)$), should be satisfied. The count of unrest events affect the audience's attention and engagement with state-run media content only to the extent that the state-run media publicizes the content to reflect the development; namely, by virtue of selecting topics for reportage, the government sets the agenda and channels public's attention. To further establish intuition; hypothetically, if the state-run media ignores the unrest event and does not modify posting behavior, the audience's engagement with government posts would not be influenced. Therefore, the mechanism through which unrest activity affects public engagement with state-run media posts conceptually relies on state-run media posts themselves. While the public may gain awareness of unrest through private, independent media conduits, this should not first-order affect the audience's engagement with the government media posts, complying with the exclusion restriction.

$$2SLS \text{ Reduced Form: } \log(Aggregate \ Attention_t) = \beta_0 + \beta_1 Count \ of \ Unrest \ Events_t + \gamma \mathbf{W}_t + \epsilon_t \quad (30)$$

$$\text{2SLS Second Stage: } \log(\text{Aggregate Attention}_t) = \beta_0 + \beta_1 \text{Content} \hat{\text{Fraction}}_t + \gamma \mathbf{W}_t + \epsilon_t \text{ where } \beta_1 = \frac{\alpha_1}{\pi_1} \quad (31)$$

4.3 Potential Limitations

Our above conceptual framework relies on the assumption that unrest events are politically destabilizing and costly for the regime. This potential for destabilization motivates the government to divert attention towards irrelevant topics. One may cite that most unrest protests in China are nonpolitical, but rather center around economic claims or other issues (e.g. demanding a nearby air polluting factory to shut down or migrate locations). However, this need not be a concern since most unrest events are inherently detrimental to social cohesion and stability even if the subject or theme is nonpolitical. Another limitation concerns audience engagement with government posts, which is subject to government manipulation. A potential direction for future research involves leveraging computer science algorithms to separate humans from “bots” deployed by the government.

5 Results and Discussion

Following days with higher unrest activity, we observe the following state-run media behavior. First, the government statistically significantly decreases the share of negative domestic political posts on Weibo, instead replacing it with neutral domestic political news. Second, the decrease in negative domestic political coverage is accompanied by a significant decrease in coverage of society, law, crime, and justice, and disaster and accident, and a significant increase in coverage of the economy. Third, overall coverage of political topics increases while coverage of nonpolitical topics decreases. Fourth, posting frequency as measured by the number of daily posts increases following unrest. Lastly, audience attention and engagement with the state-run media accounts also increase allowing government more power and flexibility to control the narrative; however, audience engagement with the negative domestic political subcategory of news decreases, to the contrary. These results collectively answer our two overarching research questions. First, state-run media indeed modifies its posting behavior, supposedly in an attempt to divert audience attention.

Specifically, it does so not by crudely decreasing political coverage and increasing nonpolitical coverage, but rather subtly by avoiding discussion of negative events and injecting neutral political information to diffuse attention and dilute negativity. Second, this diversion is potentially effective in diverting audience attention away from negative domestic political content.

Taken together, these observations suggest that the government is very subtle and sophisticated in its information management regime; it foregoes the obvious and detectable tactic of diverting attention away from politics towards entertainment, and instead pursues the nuanced maneuver of increasing neutrality of political news, while simultaneously holding positivity constant and decreasing negativity of political news. This media management tactic suggests that the government is sophisticated in anticipating a certain degree of sophistication for the individual in terms of media consumption: consider a naive citizen who cannot recognize diversion and will attend to certain topics as nudged, then the elasticity between attention and diversion is high; thus, if the government assumed the public was naive, it would use the primitive device of diversion through entertainment, namely insert entertainment content and adjust content composition to de-emphasize politics. By contrast, consider a sophisticated citizen who can detect the presence of diversion and will resist nudges for diversion, then the elasticity between attention and diversion is low; thus, if the government operated on this assumption, it would adopt the intricate technology of diversion through political neutrality and irrelevance. This empirical finding points towards level-k theory as a potentially helpful ingredient for an enriched model.

In addition to corroborating basic conjectures of the model, the empirics illuminate two interesting nuances. First, the cost of diversion and distortion on trust is high. In the model, manipulation diminishes public trust in the credibility of state-run media as an information source; this imposes a constraint that prevents the government from engaging in distortion and diversion in an unencumbered fashion. In the empirics, we observe that in the face of unrest events, state-run media actually increases content share of both domestic topics and political topics; by contrast, the mechanism through which the state diverts attention is political neutrality, not entertainment. This discrepancy suggests that the trust cost of intervention is high enough to deter the government from blatantly injecting diversion. Second, it is difficult to precisely disentangle diversion from distortion when the subject in question that the government and individual care about can only be indirectly inferred and not directly observed; i.e. x in the model may or may not be about domestic political news.

Concretely, while our operating assumption is that the state of domestic political news factors into individual protest decisions and government information manipulation decisions, an alternative possibility is that the quantity of interest is the general state of domestic news, not specifically domestic political news, that determines social unrest. This can make the two channels distortion and diversion difficult to distinguish.

Below we discuss select empirical findings.

5.1 Does the Chinese Government respond to Unrest by Engaging in Diversionary Media?

We specifically examine 11 content composition outcome variables of interest. We discuss what would theoretically precipitate in five scenarios: if the government 1) ignored unrest, 2) distorted the truth in the positive direction, 3) diverted attention towards nonpolitical topics, 4) diverted attention towards foreign topics, and 5) diverted attention towards politically neutral topics.

In Scenario 1, the government collapses all reporting about the newly emerged incident, suggesting the pre-existing distribution of news should remain constant, including the distribution of political news across sentiment categories. Therefore, the coefficient for *Fraction Negative/Neutral/Positive Given Domestic Political* should all be economically proximate to zero, similarly for the remaining coefficients. In Scenario 2, the government distorts factual truth and shifts political signals upward, suggesting a negative coefficient for *Fraction Negative Given Domestic Political* and positive coefficient for *Fraction Positive Given Domestic Political*, while being silent on the coefficient for *Fraction Neutral Given Domestic Political*. In Scenario 3, the government inserts nonpolitical content to divert attention, suggesting a negative coefficient for *Fraction Political* in Column 7, while being silent on signs for specific sentiment classes. In Scenario 4, the government inserts foreign content to divert attention, suggesting negative coefficients for *Fraction Domestic* in Column 9 and *Fraction Domestic Political* in Column 5. In Scenario 5, the government inserts mundane and irrelevant political content that is routine and uninformative about the true state of the world in order to divert attention, suggesting a negative coefficient for *Fraction Negative Given Domestic Political*, and a positive coefficient for *Fraction Neutral Given Domestic Political*.

Empirically, we find support for the fifth channel. Across all regressions, we observe an increase in neutrality and a decrease in negativity. As shown in Table 9, an additional material conflict is

contemporaneously associated with a 0.045 percent point decrease in fraction of domestic political content that is negative, a 0.058 percentage point increase in fraction of domestic political content that is neutral, and a 0.012 percentage point decrease in the fraction of domestic political content that is positive, all at statistically significant levels. Verbal conflicts are associated with similar decreases in negative domestic political news and larger increases in neutral domestic political news. Consider unrest events with nongovernmental actors, one additional event is contemporaneously associated with a 0.022 percent point decrease in fraction negative DP news, a 0.026 percentage point increase in fraction of neutral DP news, and a 0.004 percentage point decrease in the fraction of positive DP news.

Collectively, this compositional shift in sentiment suggests a rise in neutrality of news following an increase in negativity of state of affairs. This subtle tactic stands in contrast to scenarios 1 (ignore) and 2 (distort) outlined above, in which the sentiment composition either doesn't change ($\mu_d = 0$) or shifts upward ($\mu_d > 0$). Here, through the political neutrality diversion channel, the government inserts neutral and routine political events that are uninformative about the state of the world. These neutral events exist in abundance and usually happen in the background without much public attention; they include long and detailed accounts about routine party meetings, unrevealing speeches by government officials that follow conventional templates, or full length text of regulatory policies that do not pertain to the broader public. At all times, the government possesses a repository of routine meetings, speeches, and policies that state-run media can report on should it choose to. During normal times with minimal unrest activity, there is little need for state-run media to report on this; however, as positive news becomes scarce and negative news becomes abundant, this reservoir of neutrality becomes useful following unrest.

We also observe in Table 9 a statistically significant increase in the number of posts per day following unrest: an additional material conflict and an additional verbal conflict are respectively associated with 0.672 and 0.24 more posts per day from Chinese state-run media. This increase in total posts suggests evidence against Scenario 1 (ignore), as state-run media increases total reporting as a response to outbursts of unrest, instead of holding reporting constant.

Table 8 shows how the response varies over lags. Specifically, we observe that for the fraction of negative domestic political posts, the decrease is more gradual, with lagged coefficients featuring larger magnitudes than contemporaneous coefficients, suggesting that the reduction in negativity

is gradual. For the fraction of neutral domestic political posts, the increase is immediate, with contemporaneous coefficients featuring both higher statistical significance as well as magnitude, suggesting that the rise in neutrality is swift.

	Fraction Given Political	Negative Domestic	Fraction Given Political	Neutral Domestic	Fraction Given Political	Positive Domestic	Fraction Political	Fraction Domestic	Number of Posts	Log Attention Potential	Aggregate
Count Material Conflict		-0.119 (0.195)		0.529** (0.212)		-0.410* (0.227)	0.452*** (0.157)	0.143** (0.060)	0.497*** (0.186)		3.249*** (1.176)
Count Material Conflict Lag 1		-0.185 (0.183)		0.110 (0.211)		0.075 (0.218)	0.237 (0.160)	0.019 (0.063)	0.346* (0.188)		1.940* (1.175)
Count Material Conflict Lag 2		-0.376** (0.183)		-0.067 (0.189)		0.443** (0.206)	-0.185 (0.149)	-0.053 (0.060)	-0.118 (0.182)		-0.078 (1.126)
Count Verbal Conflict	-0.293 (0.183)		0.757*** (0.202)		-0.463** (0.224)		0.712*** (0.160)	0.029 (0.064)	0.297 (0.195)	1.764 (1.197)	
Count Verbal Conflict Lag 1	-0.023 (0.186)		0.100 (0.199)		-0.077 (0.220)		0.263 (0.161)	0.001 (0.063)	0.205 (0.191)	0.745 (1.131)	
Count Verbal Conflict Lag 2	-0.364** (0.174)		-0.036 (0.189)		0.401* (0.208)		-0.049 (0.154)	-0.044 (0.059)	-0.372** (0.188)	-1.510 (1.144)	
Mean Average Tone	-6.947*** (1.655)	-6.899*** (1.653)	2.457 (1.813)	1.986 (1.827)	4.490*** (1.718)	4.913*** (1.748)	4.708*** (1.251)	3.932*** (1.245)	0.865** (0.429)	-0.727 (1.136)	0.358 (1.227)
Mean Goldstein Scale	1.778 (1.375)	1.495 (1.398)	2.365* (1.296)	2.836** (1.316)	-4.143*** (1.366)	-4.332*** (1.399)	-1.195 (1.003)	-0.882 (1.034)	0.564 (0.358)	3.627*** (0.372)	4.587*** (1.138)
Day of Week	1.681 (1.177)	2.037* (1.159)	-0.865 (1.207)	-1.272 (1.166)	-0.816 (1.304)	-0.765 (1.275)	-4.874*** (0.956)	-5.102*** (0.947)	-1.273*** (0.357)	-1.041*** (0.351)	-8.355*** (0.915)
Days Linear	-0.098*** (0.003)	-0.101*** (0.003)	0.037*** (0.004)	0.043*** (0.003)	0.060*** (0.004)	0.058*** (0.003)	0.023*** (0.003)	0.030*** (0.003)	0.021*** (0.001)	0.020*** (0.001)	0.093*** (0.003)
Month	-0.458 (0.509)	-0.527 (0.511)	-2.896*** (0.537)	-2.861*** (0.540)	3.354*** (0.591)	3.388*** (0.593)	0.321 (0.460)	0.344 (0.466)	-1.171*** (0.167)	-1.157*** (0.166)	-3.189*** (0.497)
const	326.881*** (7.733)	327.032*** (7.653)	419.243*** (7.231)	421.584*** (7.088)	253.876*** (7.719)	251.384*** (7.551)	254.226*** (6.108)	873.131*** (6.092)	872.115*** (2.212)	253.439*** (2.209)	251.491*** (5.511)
Observations	2.258	2.258	2.258	2.258	2.258	2.258	2.258	2.258	2.258	2.258	2.258
R ²	0.385	0.385	0.130	0.124	0.153	0.153	0.125	0.111	0.212	0.214	0.424
Adjusted R ²	0.383	0.383	0.127	0.121	0.150	0.150	0.122	0.108	0.209	0.212	0.428
Residual Std. Error	86.694	86.692	87.659	87.988	95.477	95.494	73.830	74.395	27.044	26.999	76.568
F Statistic	153.509***	152.689***	46.456***	44.135***	49.800***	49.410***	41.562***	36.306***	70.912***	71.125***	134.745***
											142.992***
											141.114***
											152.478***

Note:

*p<0.1; **p<0.05; ***p<0.01
Note

Table 8: Preliminary OLS Fixed Effects Regression

5.1.1 By Broad Event Type

Table 9 displays summary coefficients for an array of regressors across various event types. We only report the coefficient on contemporaneous unrest for sake of brevity. Across all measures of unrest, we observe that higher counts of unrest events are associated with lower fractions of state-run media domestic political posts that are negative in sentiment, as all Column 1 coefficients are negative and statistically significant. Comparing Column 1 coefficients for Rows 4 (*Count Nongov Hostile*.) and 5 (*Count Gov Hostile*), we observe that while both coefficients are negative; the coefficient for nongovernmental hostile actions is slightly larger in magnitude than the governmental counterpart, suggesting that state-run media is less willing to highlight actions that reflect hostility from private entities, as opposed to from the government.

We also observe interesting patterns across the hostility spectrum. As shown in Column 1, reporting of negative domestic political (NDP) news decreases for both conciliatory and hostile acts, suggesting that regardless of the nature of events, state-run media's first order objective is

Regressor	Fraction Negative Given Domestic Political	Fraction Neutral Given Domestic Political	Fraction Positive Given Domestic Political	Fraction Domestic Given Political	Fraction Domestic Political	Fraction Negative Domestic Political	Fraction Political	Fraction Domestic	Number of Posts	Log Aggregate Attention Potential
Count Material Conflict	-0.455** (0.15)	0.579*** (0.152)	-0.124 (0.165)	0.466*** (0.095)	0.566*** (0.122)	0.045 (0.042)	0.515*** (0.129)	0.129*** (0.047)	0.672*** (0.132)	4.556*** (0.765)
Count Verbal Conflict	-0.5** (0.143)	0.814*** (0.145)	-0.313** (0.158)	0.248*** (0.091)	0.826*** (0.116)	0.058 (0.04)	0.868*** (0.122)	0.008 (0.045)	0.248* (0.127)	1.518** (0.737)
Count Material Cooperation	-2.222** (0.339)	1.466*** (0.346)	0.757** (0.376)	0.79*** (0.217)	1.674*** (0.276)	-0.164* (0.095)	1.643*** (0.292)	0.129 (0.106)	0.6** (0.302)	3.747** (1.756)
Count Nongov Nonviolent Hostile	-0.518** (0.151)	0.734*** (0.153)	-0.216 (0.166)	0.307*** (0.096)	0.68*** (0.122)	0.045 (0.042)	0.689*** (0.129)	0.054 (0.047)	0.355*** (0.133)	2.462*** (0.775)
Count Gov Nonviolent Hostile	-0.305** (0.072)	0.386*** (0.073)	-0.081 (0.079)	0.195*** (0.046)	0.385*** (0.058)	0.018 (0.02)	0.383*** (0.061)	0.036 (0.022)	0.244*** (0.063)	1.608*** (0.369)
Count Nongov Conciliatory	-4.457** (1.895)	2.35 (1.925)	2.107 (2.085)	-0.49 (1.204)	-1.082 (1.542)	-0.934* (0.528)	-0.995 (1.628)	0.196 (0.59)	4.851*** (1.669)	38.211*** (9.7)
Count Gov Conciliatory	-4.791*** (1.001)	3.366*** (1.018)	1.425 (1.105)	1.893*** (0.637)	5.333*** (0.81)	-0.046 (0.28)	5.365*** (0.856)	0.292 (0.313)	-0.056 (0.887)	-1.772 (5.161)

Table 9: Summary Coefficients by Event Type, Across Various Content Shares

to eliminate articles that reflect unfavorably upon the political landscape. However, the causes for decrease are different: in the case of confrontational and threatening events, the decrease in fraction NDP is motivated by disguise and diversion, whereas in the case of conciliatory events, the decrease is motivated by a genuine positive shift in affairs. This is reflected in Column 3 coefficients' signs as well as Column 1 coefficients' magnitudes. Controlling for covariates, for every increase in hostile actions, contemporaneous fraction NDP falls by 0.3-0.5 percentage points, while for every increase in conciliatory events, fraction NDP falls by 4.5-4.8 percentage points.

5.1.2 By Non-government Actor Action Type

Table 10 displays summary coefficients for various action types of nongovernmental actors and reveals interesting heterogeneity. Although coefficients are negative across all action types, they are statistically significant for conciliatory events as well as nonviolent hostile events, and not for nonviolent hostile events. suggesting two candidate explanations. A first possibility is that state-run media may report the violent hostile events (corresponding to a smaller decrease in NDP) to deter similar behavior. A second possibility is that while violent hostile acts like military posture, coercion, and assault are more physically dangerous, they are difficult to organize and implement without appropriate planning and training; to the contrary, it is precisely the nonviolent hostile acts like demands, threats, and protests that have lower activation and participation barriers.

Regressor	Fraction Negative Given Domestic Political	Fraction Neutral Given Domestic Political	Fraction Positive Given Domestic Political	Fraction Domestic Given Political	Fraction Domestic Political	Fraction Negative Domestic Political	Fraction Political	Fraction Domestic	Number of Posts	Log Aggregate Attention Potential
Total Nongov Events	-0.22*** (0.051)	0.26*** (0.052)	-0.04 (0.057)	0.141*** (0.033)	0.239*** (0.042)	0.008 (0.014)	0.232*** (0.044)	0.027* (0.016)	0.167*** (0.045)	1.136*** (0.264)
Count Nongov Ease	-4.457** (1.895)	2.35 (1.925)	2.107 (2.085)	-0.49 (1.204)	-1.082 (1.542)	-0.934* (0.528)	-0.995 (1.628)	0.196 (0.59)	4.851*** (1.669)	38.211*** (9.7)
Count Nongov Demand	-3.722*** (1.036)	2.268** (1.053)	1.454 (1.142)	1.536** (0.659)	2.914*** (0.842)	-0.2 (0.29)	2.878*** (0.89)	0.304 (0.323)	1.548* (0.915)	11.713** (5.326)
Count Nongov Disapprove	-0.722** (0.307)	1.4*** (0.311)	-0.677** (0.338)	0.431** (0.195)	0.942*** (0.249)	0.099 (0.086)	0.973*** (0.263)	-0.057 (0.096)	0.444 (0.271)	2.721* (1.578)
Count Nongov Threat	-1.721** (0.735)	3.062*** (0.744)	-1.342* (0.808)	0.739 (0.467)	2.594*** (0.596)	0.045 (0.205)	2.711*** (0.629)	0.221 (0.229)	0.791 (0.648)	5.242 (3.774)
Count Nongov Protest	-3.941*** (1.032)	2.263** (1.049)	1.678 (1.137)	1.283* (0.656)	4.379*** (0.836)	-0.277 (0.288)	4.484*** (0.883)	0.117 (0.322)	-0.713 (0.912)	-3.997 (5.309)
Count Nongov Posture	-2.137 (1.424)	-1.165 (1.445)	3.302** (1.564)	1.685* (0.903)	-1.701 (1.157)	-0.656* (0.397)	-2.403** (1.221)	0.256 (0.443)	-5.652*** (1.25)	- (7.288)
Count Nongov Coerce	-0.786 (1.56)	2.147 (1.583)	-1.361 (1.714)	0.939 (0.99)	-0.547 (1.268)	-0.234 (0.435)	-0.723 (1.339)	-0.572 (0.485)	-4.748*** (1.371)	- (7.986)
Count Nongov Assault	-0.797 (0.911)	1.188 (0.924)	-0.391 (1.001)	0.995* (0.578)	2.175*** (0.739)	0.393 (0.254)	2.259*** (0.78)	0.183 (0.283)	-2.546*** (0.801)	- (4.658)
Count Nongov Fight	-0.603 (0.512)	1.468*** (0.519)	-0.865 (0.562)	0.638** (0.325)	1.714*** (0.415)	0.318** (0.142)	1.746*** (0.438)	0.54*** (0.159)	4.287*** (0.442)	27.891*** (2.56)

Table 10: Summary Coefficients by Action Type, for Nongov Actors, Across Various Content Shares

Furthermore, while violent hostile acts are usually isolated and small-scale, nonviolent hostile acts have substantial potential to be collective and large-scale, allowing for collective mass mobilization on a societal level, especially through online social media platforms. Therefore, the mobilization potential that is present in nonviolent hostility but absent in violent hostility points towards nonviolent hostility as the government's bigger fear; accordingly, to prevent inspiring future protests, state-run media selectively restricts reporting.

Regressor	Fraction Negative Given Domestic Political	Fraction Neutral Given Domestic Political	Fraction Positive Given Domestic Political	Fraction Domestic Given Political	Fraction Domestic Political	Fraction Negative Domestic Political	Fraction Political	Fraction Domestic	Number of Posts	Log Aggregate Attention Potential
Total Nongov Events	-0.22*** (0.051)	0.26*** (0.052)	-0.04 (0.057)	0.141*** (0.033)	0.239*** (0.042)	0.008 (0.014)	0.232*** (0.044)	0.027* (0.016)	0.167*** (0.045)	1.136*** (0.264)
Count Nongov Political Issue	-1.245 (1.479)	4.03*** (1.499)	-2.785* (1.625)	0.791 (0.939)	0.12 (1.203)	-0.429 (0.412)	-0.01 (1.27)	-0.685 (0.46)	-4.244*** (1.301)	- (7.573)
Count Nongov Economic Issue	-0.944 (2.118)	1.132 (2.149)	-0.189 (2.328)	3.351** (1.342)	2.667 (1.721)	0.709 (0.59)	2.058 (1.817)	0.661 (0.658)	-4.421** (1.864)	- (10.842)
Count Nongov Military Issue	-0.742** (0.343)	0.959*** (0.348)	-0.217 (0.378)	0.513** (0.218)	1.043*** (0.278)	0.127 (0.096)	1.041*** (0.294)	0.291*** (0.107)	1.309*** (0.302)	9.352*** (1.752)

Table 11: Summary Coefficients by Topic Type, for Nongov Actors, Across Various Content Shares

5.1.3 By Target of Nongov Action

Table 12 shows summary coefficients for government vs. non government targets of non government entity actions. Interestingly, state-run media behavior is starkly different for different types of targets. For actions that are dangerous (e.g. protests and threats) or highly dangerous (assaults and fights), pairwise comparisons of coefficients for “nongov onto gov” and “nongov onto nongov” reveals that state-run media suppresses coverage when these actions threaten government institutions, but not when the target entity is nongovernmental. This phenomenon manifests both in coefficients and in statistical significance: the fraction of domestic political posts that have negative sentiment decreases by 0.132 percentage points at significance level of 0.01 for every increase in unrest events where the victim of nongovernmental action is the government, whereas the corresponding decrease is much smaller at 0.33 percentage points non statistically significantly for events where the victim is a private entity. Namely, the government allows news that portray hazardous behavior between private entities, but suppresses news of private entities inflicting harm upon the government, likely because the former does not implicate the government action and is devoid of political destabilization risk, while the latter jeopardizes the image and authority of the government and can spark collective mobilization. In the interest of preventing the spread of rumors and speculation that heightens unrest, the government aims to curtail collective action by selectively under-reporting events that portray civilians challenging the government’s supremacy legal ordinances. This is congruent with King, Pan and Roberts (2017) findings that the Chinese government is more concerned with posts that catalyze collective coordination, compared to criticisms that carry no call to action. To the contrary, when unrest exclusively involves private entities, it is viewed as less threatening to state control. In some cases, the government may even allow coverage of such events to deflect attention away from more sensitive issues that impair the regime’s authority. Total post counts and log aggregate attention potential unveil similar propensities. State-run media are more likely to increase posting and garner greater audience engagement when the target in question is a private entity, manifested in the statistical significance.

Regressor	Fraction Negative Given Domestic Political	Fraction Neutral Given Domestic Political	Fraction Positive Given Domestic Political	Fraction Domestic Given Political	Fraction Domestic Political	Fraction Negative Domestic Political	Fraction Political	Fraction Domestic	Number of Posts	Log Aggregate Attention Potential
Count Nongov Violent and Hostile to Gov	-1.327** (0.612)	1.602*** (0.621)	-0.274 (0.674)	1.091*** (0.388)	1.785*** (0.497)	0.111 (0.171)	1.705*** (0.525)	0.367* (0.19)	0.848 (0.54)	6.138* (3.142)
Count Nongov Violent and Hostile to Nongov	-0.33 (0.547)	0.92* (0.555)	-0.59 (0.601)	0.581* (0.347)	1.001** (0.444)	0.291* (0.152)	1.007** (0.469)	0.358** (0.17)	1.884*** (0.481)	13.596*** (2.792)
Count Nongov Nonviolent and Hostile to Gov	-0.836*** (0.294)	1.723*** (0.296)	-0.888*** (0.323)	0.549*** (0.186)	1.695*** (0.237)	0.158* (0.082)	1.758*** (0.25)	0.055 (0.091)	0.597** (0.259)	3.236** (1.509)
Count Nongov Nonviolent and Hostile to Nongov	-0.745*** (0.24)	0.697*** (0.243)	0.047 (0.264)	0.406*** (0.152)	0.58*** (0.195)	0.008 (0.067)	0.561*** (0.206)	0.1 (0.075)	0.495** (0.211)	4.033*** (1.229)
Count Nongov Conciliatory to Gov	-2.459 (3.278)	1.157 (3.326)	1.302 (3.602)	-0.26 (2.081)	-3.193 (2.663)	-0.871 (0.913)	-3.167 (2.812)	0.492 (1.019)	-4.172 (2.888)	-14.054 (16.814)
Count Nongov Conciliatory to Nongov	-6.132** (2.462)	3.312 (2.5)	2.82 (2.708)	-0.681 (1.564)	-0.021 (2.003)	-1.084 (0.686)	0.112 (2.115)	0.052 (0.766)	10.546*** (2.161)	72.441*** (12.553)

Table 12: Summary Coefficients by Action and Target Type, for Nongov Actors, Across Various Content Shares

5.1.4 By Multiclass Topic Classification

Table 13 illustrates how state-run media shifts the topic composition. As verbal and material conflicts increase, state-run media proportionally increases the number of articles about the economy: with the appropriate time controls, and controlling for goldstein scale and average tone, an additional verbal and material conflict increases the economy proportion by 0.1688%, at a 0.01 confidence level.

The government may increase economic posts following unrest for three main reasons: 1) to demonstrate government competence, 2) to reassure the public, and 3) to generally divert attention. First, during times of unrest, the government may be perceived as ineffective or incompetent. By highlighting positive economic news and demonstrating sound economic management, state-run media can try to restore people’s confidence in the party leadership. Second, the down-stream ripple effects of political and social unrest may engender public fears about the ominous state of the economy; amplifying the visibility of economic content can provide a placating effect of alleviating concerns.

Simultaneously, we observe a decrease in posts about law, crime, and justice, as well as disaster and accident, for reasons of 1) avoiding attracting interest and inciting further unrest, and 2) maintaining state control of narrative. First, if the event was related to issues of law or justice, sustaining coverage on these topics could inflame tensions and escalate the tumult; therefore decreasing coverage on contentious topics decreases helps to conclude the unrest episode. Second, since the state-run media is an important instrument for controlling the narrative, by posting less

about law, crime, and justice, the government can steer public discourse in a favorable direction and avoid undesirable unforeseen consequences. Similarly, societal posts also decrease in proportion. Other topics do not exhibit notable patterns.

	Fraction Economy	Fraction Law, Crime, Justice	Fraction Disaster and Accident	Fraction Health	Fraction Culture	Fraction Society	Fraction Lifestyle Leisure	Fraction Environment	Fraction Education	Fraction Science Technology Engineering Math	Fraction Sport
Count Verbal Material Conflict	1.688*** (0.588)	-2.212*** (0.589)	-1.367*** (0.501)	0.748 (0.988)	-0.347 (0.689)	-0.533 (0.695)	-0.261 (0.769)	0.354 (0.467)	0.469 (0.332)	-0.554 (0.353)	-1.578*** (0.448)
Day of Week	-6.142*** (0.541)	-2.697*** (0.542)	2.362*** (0.461)	-0.276 (0.910)	1.599** (0.634)	1.908*** (0.640)	3.309*** (0.708)	1.273*** (0.430)	-0.937*** (0.306)	0.924*** (0.325)	2.088*** (0.412)
Days Linear	-0.044*** (0.002)	-0.080*** (0.002)	-0.014*** (0.001)	0.042*** (0.003)	0.023*** (0.002)	0.026*** (0.002)	-0.055*** (0.002)	0.007*** (0.001)	-0.009*** (0.001)	0.005*** (0.001)	-0.004*** (0.001)
Mean Average Tone	1.843** (0.790)	-2.828*** (0.791)	-1.575** (0.673)	-5.373*** (1.327)	1.509 (0.925)	2.156** (0.933)	1.011 (1.033)	0.278 (0.627)	1.122** (0.446)	1.106** (0.474)	-1.093* (0.601)
Mean Goldstein Scale	0.301 (0.666)	2.584*** (0.667)	0.307 (0.568)	2.936*** (1.120)	-0.878 (0.781)	-0.274 (0.787)	-0.986 (0.872)	-1.541*** (0.529)	0.156 (0.376)	-1.173*** (0.400)	0.561 (0.507)
Month	0.064 (0.311)	1.638*** (0.312)	0.851*** (0.265)	-6.025*** (0.523)	0.123 (0.365)	0.218 (0.368)	-0.076 (0.407)	0.800*** (0.247)	0.931*** (0.176)	1.319*** (0.187)	1.013*** (0.237)
const	342.973*** (4.067)	305.394*** (4.074)	102.540*** (3.465)	84.826*** (6.835)	160.105*** (4.766)	297.541*** (4.807)	389.004*** (5.323)	136.599*** (3.231)	84.788*** (2.297)	76.199*** (2.441)	63.313*** (3.098)
Observations	2,260	2,260	2,260	2,260	2,260	2,260	2,260	2,260	2,260	2,260	2,260
R^2	0.275	0.531	0.064	0.158	0.060	0.075	0.238	0.025	0.061	0.039	0.034
Adjusted R^2	0.273	0.530	0.062	0.155	0.058	0.072	0.236	0.023	0.059	0.036	0.032
Residual Std. Error	50.810	50.898	43.292	85.389	59.546	60.063	66.497	40.366	28.693	30.502	38.703
F Statistic	142.547***	424.762***	25.713***	70.226***	24.164***	30.365***	116.986***	9.706***	24.578***	15.257***	13.350***

Note:

*p<0.1; **p<0.05; ***p<0.01
Note

Table 13: Summary Coefficients across Weibo Post Content Categories

5.1.5 By Destabilization Level

Table 14 offers a granular decomposition by segmenting events by destabilization potential defined above, and separately regressing media intervention variables for each subset of events. Separating destabilization potentials into quartiles, we observe that non-reporting of negative domestic political news is more pronounced for more destabilizing events. The coefficient decrease in magnitude in the negative direction and increase in statistical significance as the destabilization quartile increases. While counts of verbal material conflict are actually associated with a 0.027 percentage point *increase* in fraction of NDP news for 25% least destabilizing unrest events (albeit statistically insignificantly), they are associated with a 0.25 percentage point decrease in fraction NDP for the 25% most destabilizing unrest events, with statistical significance.

	<i>Fraction Negative Given Domestic Political</i>			
	<i>First Quartile</i>	<i>Second Quartile</i>	<i>Third Quartile</i>	<i>Fourth Quartile</i>
Count Verbal Material Conflict	0.270 (0.205)	-0.291 (0.230)	-1.092*** (0.238)	-2.507*** (0.230)
Mean Goldstein Scale	2.249** (1.133)	1.990* (1.134)	1.652 (1.128)	2.221** (1.100)
Mean Average Tone	-4.948*** (1.466)	-6.119*** (1.380)	-6.362*** (1.345)	-5.637*** (1.310)
Days Linear	-0.107*** (0.003)	-0.104*** (0.003)	-0.099*** (0.003)	-0.092*** (0.003)
Day of Week	1.767* (0.945)	1.125 (0.951)	0.234 (0.949)	-1.679* (0.937)
Month	-0.456 (0.532)	-0.495 (0.533)	-0.559 (0.530)	-0.473 (0.519)
Observations	2,260	2,260	2,260	2,260
R^2	0.382	0.382	0.387	0.413
Adjusted R^2	0.380	0.380	0.386	0.411
Residual Std. Error	86.943	86.946	86.574	84.766
F Statistic	232.232***	232.198***	237.428***	263.845***

Note:

*p<0.1; **p<0.05; ***p<0.01
Note

Table 14: OLS Fixed Effects Regression by Quartiles of Destabilization Potential

For the instrumental variables two stage least squares approach, first stage regressions shown in Table 15 that each instrument is individually strong for the treatment “Count of Nongovernment Hostile Events,” as shown by F statistics that clear the Montiel-Pflueger critical threshold of 23.1. Second stage regressions reveal that the choice of instruments causes perturbations in the

magnitudes of the estimate. This suggests that these instruments could be estimating different local average treatment effects. For instance, since the instrument *Max Temp* and *Gust* varies across days in different patterns, the local subsample of days that are subject to treatment effect from these instruments may be different, explaining the variance in coefficients. Nonetheless, overall across instruments, higher counts of nongovernmental hostile events are associated with lower coverage of negative domestic political events.

<i>Count of Nongovernment Hostile Events</i>													
Thunder	1.83**												
	(0.66)												
Hail		1.61											
		(2.04)											
Max Temp			0.03**										
			(0.01)										
Min Temp				0.03**									
				(0.01)									
Rain					0.78*								
					(0.32)								
Snow						1.18*							
						(0.60)							
Thunder x Maxtemp							0.02**						
							(0.01)						
Thunder x Gust								0.07**					
								(0.02)					
Thunder x Hail									1.61				
									(2.04)				
Hail x Gustspeed										0.17**			
										(0.06)			
Maxtemp High											1.69**		
											(0.62)		
Hail x Maxtemp												0.02	
												(0.02)	
Hail x Rain													1.61
													(2.04)
Constant	4.56**	4.56**	4.67**	3.02**	3.42**	4.52**	4.63**	4.63**	4.67**	4.67**	4.67**	4.67**	4.67**
	(0.33)	(0.33)	(0.32)	(0.48)	(0.39)	(0.33)	(0.33)	(0.33)	(0.32)	(0.32)	(0.32)	(0.32)	(0.32)
Obs	2218	2218	2218	2218	2218	2218	2218	2188	2218	2218	2244	2218	2218
Adj. R ²	0.20	0.20	0.19	0.20	0.21	0.20	0.20	0.20	0.20	0.19	0.20	0.19	0.19
F-stat	516.98	521.44	493.03	496.13	499.23	507.44	505.17	491.72	507.58	493.03	501.15	494.47	493.03

Note: This table presents the first stage regression for IV2SLS. The dependent variable is the count of nongovernmental actor hostile events. Controls like month and linear time trend are included to make the instrument as good as randomly assigned.

Table 15: Weather Instrument 2SLS First Stage Regression

5.2 Does government diversionary media effectively divert public attention?

Collectively, OLS results suggest that the state-run media's strategy of altering its content profile is highly effective for capturing attention. Specifically, this occurs in two simultaneous channels. First, Table 17 suggests that state-run media are successful in attracting netizens to their accounts, and in so doing prevails in the competition against other media outlets for audience attention. This allows government media to dominate if not monopolize the public narrative, and as such can offer an innocuous account less sinister than the free media alternative. Second, Table 18, in concert

Dependent Variable Instrument	<i>Fraction Negative Given Domestic Political</i>													
	Thunder	Hail	Max Temp	Min Temp	Rain	Thunder x Maxtemp	Thunder x Gust	Thunder x Hail	Hail x Gustspeed	x Maxtemp High	Hail x Maxtemp	x Rain	Hail x	
Mean Goldstein Scale	6.15 (7.45)	-37.89 (52.52)	6.08 (7.99)	3.12 (6.62)	6.84 (11.63)	6.73 (7.48)	8.83 (8.62)	-37.89 (52.52)	-17.60 (11.82)	4.80 (9.51)	-30.91 (34.93)	-37.89 (52.52)		
Mean Average Tone	2.01 (8.12)	-43.12 (54.15)	1.95 (8.65)	-1.08 (7.48)	2.73 (12.69)	2.61 (8.16)	4.89 (9.13)	-43.12 (54.15)	-22.33+ (12.94)	2.79 (10.52)	-35.97 (36.21)	-43.12 (54.15)		
Day of Week	2.54 (7.40)	-42.06 (53.10)	2.48 (8.13)	-0.52 (6.84)	3.25 (12.14)	3.13 (7.45)	5.42 (8.51)	-42.06 (53.10)	-21.51+ (11.94)	2.06 (9.49)	-34.99 (35.31)	-42.06 (53.10)		
Days Linear	-0.10** (0.03)	0.08 (0.21)	-0.10** (0.03)	-0.08** (0.03)	-0.10* (0.05)	-0.10** (0.03)	-0.11** (0.03)	0.08 (0.21)	-0.00 (0.05)	-0.09* (0.04)	0.05 (0.14)	0.08 (0.21)		
Month	2.38 (1.56)	4.08 (3.13)	2.38 (1.57)	2.49 (1.56)	2.35 (1.61)	2.36 (1.56)	2.59 (1.58)	4.08 (3.13)	3.29+ (1.85)	2.31 (1.57)	3.81 (2.54)	4.08 (3.13)		
Count Nongov Hostile Events	2.62 (10.56)	-67.01 (82.46)	2.52 (11.57)	-2.16 (9.36)	3.72 (18.31)	3.54 (10.63)	7.49 (12.37)	-67.01 (82.46)	-34.93* (17.79)	1.81 (14.12)	-55.97 (54.62)	-67.01 (82.46)		
Constant	308.86** (33.93)	482.89* (207.78)	309.11** (35.59)	320.80** (31.45)	306.10** (50.23)	306.55** (34.07)	294.42** (38.57)	482.89* (207.78)	402.72** (49.86)	311.90** (39.92)	455.31** (138.81)	482.89* (207.78)		
Obs	2089	2089	2089	2089	2089	2089	2062	2089	2089	2114	2089	2089		
Adj. R ²	0.05	-1.86	0.05	0.06	0.05	0.05	0.02	-1.86	-0.43	0.06	-1.26	-1.86		
F-stat	137.76	53.06	136.91	139.97	135.70	136.97	131.74	53.06	104.39	139.32	66.69	53.06		

Table 16: Weather Instrument IV2SLS Second Stage

with previous findings in part 1, suggests that when unrest potential is high, decreasing coverage of negative domestic political topics potentially decreases audience attention on these bleak subjects, thereby deflecting scrutiny towards other more neutral topics.

As a preliminary sense check, OLS Table 27 decomposes aggregate daily engagement by topic to confirm that, for each topic, as the proportional prevalence increases, the audience attention and engagement with that topic is amplified correspondingly. For each of the topics negative, neutral, and positive domestic political, engagement with the particular category of posts increases with its frequency, conforming with intuition.

Table 17 examines how audience attention towards all state-run media posts evolve as the content composition changes. We observe that audience engagement with state-run media is increasing in both the fraction of domestic political news that are negative and the overall fraction of political news, and decreasing in the fraction of domestic political news that are positive all in a statistically significant manner. Since we estimate a log-linear regression, a one percentage point increase in *Fraction Negative Given Domestic Political* article share is associated with an increase in the audience attention potential by 960 units, controlling for time trends and unrest characteristics, while a one percentage point increase in *Fraction Positive Given Domestic Political* article share is associated with a 772 unit decrease in attention, controlling for covariates. Candidate explanations for this phenomenon is the negativity bias, whereby negative news are more salient than positive events, capturing greater attention and recall. Our unconscious allocation of attention to negative

events explains the prevalence of sensational journalism. Likewise, we observe in the state-media context that total engagement per day is strongly increasing in the fraction of negative news and political news.

<i>Log Aggregate Attention Potential</i>						
Fraction Negative Given Domestic Political	0.960*** (0.106)					
Fraction Neutral Given Domestic Political		-0.032 (0.106)				
Fraction Positive Given Domestic Political			-0.772*** (0.097)			
Fraction Domestic Given Political				0.013 (0.170)		
Fraction Negative Domestic Political					5.731*** (0.368)	
Fraction Political						1.926*** (0.119)
Mean Goldstein Scale	19.483*** (5.598)	23.612*** (5.684)	18.829*** (5.634)	23.567*** (5.684)	19.116*** (5.405)	22.121*** (5.377)
Mean Average Tone	-1.460 (6.816)	-6.416 (6.917)	-1.666 (6.847)	-6.415 (6.928)	-1.446 (6.580)	-9.589 (6.549)
Days Linear	0.694*** (0.019)	0.594*** (0.017)	0.638*** (0.017)	0.592*** (0.017)	0.694*** (0.016)	0.525*** (0.016)
Day of Week	-32.189*** (4.607)	-30.953*** (4.694)	-30.103*** (4.624)	-30.849*** (4.704)	-24.441*** (4.474)	-18.704*** (4.500)
Month	-19.102*** (2.674)	-19.747*** (2.738)	-17.002*** (2.704)	-19.645*** (2.725)	-19.449*** (2.586)	-20.278*** (2.576)
Observations	2,260	2,260	2,260	2,260	2,260	2,260
R^2	0.456	0.436	0.451	0.436	0.491	0.495
Adjusted R^2	0.454	0.434	0.450	0.434	0.489	0.493
Residual Std. Error	437.675	445.546	439.446	445.554	423.383	421.709
F Statistic	314.246***	290.090***	308.697***	290.066***	361.597***	367.460***

Note: *p<0.1; **p<0.05; ***p<0.01
Note

Table 17: OLS Specification: Log Aggregate Attention and Engagement Against Content Compositional Shifts

Table 18 evidences that attention on negative domestic political (NDP) issues are increasing in their relative share, suggesting that netizens are easily prone to nudges. Incorporating our finding that NDP relative share decreases following unrest, we deduce that audience engagement with NDP issues indeed decreases following unrest. Namely, the audience consumes media in an unsophisticated manner and does not compensate for the decreased coverage by engaging in a higher-than-baseline way.

IV2SLS specification (Table 19) conforms with our OLS findings in terms of directions and levels but suggest lower statistical significance for certain regressors. IV2SLS indicates that the conditional proportion of negative domestic news (*Fraction Negative Given Domestic Political*) but

<i>Log Attention Negative Domestic Political</i>						
Fraction Negative Given Domestic Political	0.397*** (0.011)					
Fraction Neutral Given Domestic Political		-0.098*** (0.013)				
Fraction Positive Given Domestic Political			-0.245*** (0.011)			
Fraction Domestic Given Political				0.055** (0.021)		
Fraction Negative Domestic Political					1.964*** (0.026)	
Fraction Political						0.347*** (0.014)
Mean Goldstein Scale	0.748 (0.567)	2.597*** (0.708)	0.893 (0.655)	2.463*** (0.716)	0.911** (0.382)	2.151*** (0.636)
Mean Average Tone	-1.071 (0.690)	-3.180*** (0.861)	-1.574** (0.796)	-3.202*** (0.872)	-1.420*** (0.464)	-3.669*** (0.774)
Days Linear	0.052*** (0.002)	0.015*** (0.002)	0.025*** (0.002)	0.009*** (0.002)	0.045*** (0.001)	-0.002 (0.002)
Day of Week	-3.824*** (0.467)	-3.522*** (0.585)	-3.029*** (0.538)	-3.159*** (0.592)	-1.069*** (0.316)	-1.079** (0.532)
Month	-1.026*** (0.271)	-1.556*** (0.341)	-0.432 (0.314)	-1.233*** (0.343)	-1.177*** (0.183)	-1.374*** (0.304)
Observations	2,258	2,258	2,258	2,258	2,258	2,258
R^2	0.403	0.065	0.208	0.045	0.729	0.246
Adjusted R^2	0.402	0.063	0.206	0.043	0.728	0.244
Residual Std. Error	44.313	55.453	51.042	56.047	29.861	49.807
F Statistic	253.432***	26.239***	98.615***	17.771***	1009.070***	122.398***
<i>Note:</i>				*p<0.1; **p<0.05; ***p<0.01 Note		

Table 18: OLS Specification: Log Attention Towards Negative Domestic Political Content Against Content Compositional Shifts

<i>Log Attention Negative Domestic Political</i>						
Fraction Negative Given Domestic Political	0.58** (0.17)					
Fraction Neutral Given Domestic Political		-0.36** (0.13)				
Fraction Positive Given Domestic Political			0.97 (0.76)			
Fraction Domestic Given Political				-1.72 (1.20)		
Fraction Negative Domestic Political					11.06 (15.17)	
Fraction Political						-0.56+ (0.29)
Constant	-106.80+ (55.90)	237.54** (56.58)	-151.97 (185.57)	1584.13 (1044.37)	-742.24 (1135.40)	231.05** (75.47)
Obs	2258	2258	2258	2258	2258	2258
Adj. R ²	0.33	-0.10	-3.87	-2.88	-14.03	-1.14
F-stat	154.77	94.53	21.39	26.82	6.92	48.72

Note: This table shows IV2SLS estimation results using the instrument “Count Verbal and Material Conflict.” The regular battery of controls are included, including mean goldstein scale, mean average tone, day of week, days linear, and month. Lags in the dependent variable are not included.

Table 19: IV2SLS Reduced Form: Log Attention On Negative Domestic Political Events

not the raw proportion (*Fraction Negative Domestic Political*) matters for audience attention directed towards negative domestic political news. Additionally, while the coefficient conditional proportions of positive DP news is insignificant, the conditional proportion of neutral DP news is statistically significant. In the log-linear regression, a one percentage point increase in *Fraction Neutral Given Domestic Political* article share is associated with a decreases in the audience attention on negative domestic political topics by 360 units, controlling for covariates.

6 Conclusion

In conclusion, we find evidence that the Chinese state-run media engage in strategic diversion to deflect public attention away from negative domestic political news. Specifically, it achieves so not through diversion with entertainment, but rather through diversion with political neutrality; namely, instead of catching the public eye with engaging or sensational nonpolitical content, the state-run media instead injects neutral and mundane political information that effectively dilutes the negativity.

In terms of empirics, directions for future research are various. First, we can make cross sectional comparisons across media outlets, first between government-run and private-run media, and second between local and national level state-run media, in order to derive an elasticity interpretation. Media outlets that are more explicitly state run and politically more aligned with the CCP should exhibit lower elasticity to unrest (i.e. more anchored towards government internal benchmark objective) than publications that are less closely tied. Second, data permitting, we can broaden our IV regression to incorporate geographical dimensions and local government-run media outlets. Third, it could be interesting to study the behavior of internet “bots” hired by the government. Aside from the government distracting people through direct activity of government institution accounts, the government can also hire people to insert large numbers of pseudonymous and fabricated writings under the guise of real Weibo users. These government hired employees pretend to represent the genuine opinion of ordinary people. Many of these posts involve national cheerleading, revolutionary history of the Communist Party, and other patriotic topics. Future research can aim to detect bot behavior and examine whether bots achieve similar flooding effects as state-run media.

In terms of theory, there are two further possibilities. First, the model can be enriched with elements of limited attention (rational or behavioral), treating attention as endogenously determined by the salience of the news as well as the number of competing news. This allows further microfounding of the individual problem. Second, our parsimonious model can be further enriched with distinctions between naive and sophisticated individuals. In the current setup, only the government is sophisticated in internalizing the individual’s protest response function. In the world with sophisticated individuals, they anticipate and account for the government’s manipulation

by effectively undoing it during signal extraction. Assuming a heterogeneous population with both naive and sophisticated agents, the government’s optimal action would change depending on the aggregate population profile. Third, the current model examines how distortion and diversion operate in the domestic context, and models mechanisms of sentiment shift and topic topic shift; however, since it only concerns the home country, it cannot model the channel of location diversion (through which state-run media decrease coverage of negative domestic political events and increase coverage of negative foreign political events). If we generalize the model to redefine x not as the absolute level of unrest in the home country, but rather as the relative difference in unrest between home country and foreign country (to the extent that individuals care about how comparatively “good” their country is relative to others), we can account for the location diversion channel. Fourth, the model can be extended to model different scenarios of where unrest arises from: i.e. exogenous political shocks vs. endogenous economic shocks. This distinction implies differences in how quickly unrest precipitates and subsequently how state-run media responds.

A Appendix

B Weibo Data

Example Posts Below are example posts for each type of label:

- Political, Patriotic, Attention Grabbing:
 - **【人民日报钟声：所谓改变中国图谋纯属痴人说梦】** 一段时间以来，美国一些政客四处兜售谎言，恶意攻击中国，行尽造谣中伤之能事。无论他们怎样煞费心机自我伪装，其所谓“改变中国”的祸心早已昭然于世。中国共产党的领导和中国道路，是历史的选择，人民的选择。美国一些政客试图阻挠中国走向美好未来的任何企图，都是螳臂当车，不自量力。人民日报钟声：所谓“改变中国”图谋纯属痴人说梦
 - Translation: [People’s Daily Bell: The so-called plot to change China is purely a dream of idiots] For a period of time, some politicians in the United States have been peddling lies everywhere, maliciously attacking China, and doing everything they can to spread rumors and slander. No matter how much they try to disguise themselves, their so-called evil intention of “changing China” has long been clear to the world. The leadership of the Communist Party of China and the Chinese path are the choice of history and the choice of the people. Any attempt by some American politicians to obstruct China’s progress towards a better future is like a mantis’ arm and a cart, and they are overwhelmed. People’s Daily Bell: The so-called “change China” plot is purely a dream
- Political, Nonpatriotic, Attention Grabbing:
 - **【波兰总统和冒充马克龙博主聊了 7 分钟】** 据路透社报道，波兰总统办公室当地时间 22 日表示，15 日波兰导弹事件发生后，波兰总统杜达与一名冒充法国总统马克龙的恶作剧博主进行了电话交谈，杜达告诉这位“法国总统”，自己“不想与俄罗斯发生战争”。但不久后，杜达“从对话者不同寻常的谈话方式中”意识到这可能是一场恶作剧，于是结束了对话。波兰方面目前正在调查来电者是如何设法打通电话的。报道称，打电话给波兰总统的是俄罗斯恶作剧博主沃万和列克苏斯，这是他们时隔两年再次拨通波兰总统的电话。2020 年，他们还冒充联合国秘书长古特雷斯致电，引发外界对波兰总统办公室安全和电话甄别问题的质疑。这二人还曾骗过马克龙、英国歌手埃尔顿·约翰和英国前首相鲍里斯·约翰逊等人。波兰总统被“假马克龙”骗聊了 7 分钟，信息量很大
 - Translation:[Poland President chatted with a blogger pretending to be Macron for 7 minutes] According to Reuters, the Polish President’s Office stated on the 22nd local time that after the Polish missile incident on the 15th, Polish President Duda and a person posing as French President Macron In a phone conversation with the prank blogger, Duda told the “French president” that he “doesn’t want a war with Russia.” But shortly after, Duda realized “from the unusual manner of speaking of the interlocutor” that this might be a prank, and ended the conversation. Poland is currently investigating how the caller managed to get through. According to the report, the Russian prank bloggers Vowan and Lexus called the Polish president. This is the second time they called the Polish president after two years. In 2020, they also pretended to be the Secretary-General of the United Nations Guterres to call, raising questions about the security

of the Polish President's office and phone screening issues. The duo have also duped Macron, British singer Elton John and former British Prime Minister Boris Johnson, among others. The Polish President was tricked by the "fake Macron" for 7 minutes, with a lot of information.

- Nonpolitical, Nonpatriotic, Attention Grabbing:

- **【男子因过年没钱花绑架自己向妻子索 10 万赎金】**家里的钱都归老婆管，“没钱过年”的宜宾男子唐某竟谎称自己被绑架，向老婆索要 10 万赎金，还躺进汽车后备箱，用胶带封住自己嘴巴拍照。唐某的老婆信以为真，吓得报警，警方很快查明真相。唐某被拘留 10 日。via 成都商报男子绑架自己向妻索财十万’
- Translation: [The man kidnapped himself and demanded 100,000 ransom from his wife because he had no money to spend during the Chinese New Year] All the money in the house belonged to his wife. Tang, a Yibin man who "had no money for the Chinese New Year", lied that he was kidnapped and demanded 100,000 ransom from his wife. Tang got into the trunk of a car, taped his mouth shut, and took a photo. Tang's wife believed it was true and was so frightened that she called the police. The police quickly found out the truth. Tang was detained for 10 days. Via Chengdu Commercial Daily: Man kidnaps himself and asks wife for 100,000 yuan.

- Nonpolitical, Patriotic, Attention Grabbing:

- **【6 秒 43 再创亚洲纪录！苏炳添室内巡回赛男子 60 米夺冠】**国际田联消息，在德国杜塞尔多夫周二举行的 2018 国际田联世界室内巡回赛男子 60 米决赛中，中国选手苏炳添以 6 秒 43 的成绩打破亚洲纪录夺冠。赋害了，中国飞人！
- Translation: [6.43 seconds set another Asian record! Su Bingtian won the men's 60m championship in the Indoor Tournament] According to IAAF news, in the men's 60m final of the 2018 IAAF World Indoor Tournament in Dusseldorf, Germany on Tuesday, Chinese contestant Su Bingtian broke the Asian record with a time of 6.43 seconds and won the championship. Bravo, Chinese flying man!

- No region:

- **【夜读：有的人只活了 1 天，然后重复了 364 遍】**你决定读书，看了两页就昏昏欲睡；想健身，却从没尝试。你想过很多改变，却一次都没做到。最后又回到了以前的状态，躺在床上，拿着手机刷个剧。对那个爱拖延的自己好好说声再见吧。2017 年的最后三个月，努力把之前的半途而废都弥补上。毕竟，时间不会等你。
- Translation: [Night reading: Some people only live for 1 day, and then repeat it 364 times] You resolve to read, and you fall asleep after reading two pages; you resolve to exercise, but you never try. You have thought about making many changes, but you have not accomplished it once. In the end, you return to the previous state, lying in bed, watching a drama with your mobile phone. Say goodbye to that procrastinating self. In the last three months of 2017, try your best to make up for giving up previously. After all, time will not wait for you.

- Global:

- **【全球一半儿童存在隐性饥饿问题】**联合国儿童基金会日前发布的《2019 年世界儿童状况：儿童、食物与营养》报告显示，全球有一半儿童存在“隐性饥饿”问题。什么是“隐性饥饿”？饱了还会饿？（科技日报记者张佳星）全球一半儿童存在“隐性饥饿”问题

- Translation: [Half of the world’s children have hidden hunger problem]UNICEF recently released the ”2019 State of the World’s Children: Children, Food and Nutrition” report, showing that half of the world’s children suffer from ”hidden hunger”. What is ”hidden hunger”? Will you still be hungry? (Science and Technology Daily reporter Zhang Jiaxing) Half of the world’s children suffer from ”hidden hunger”

Media	Chinese Name	Administrative Level	Subscribers	Total Posts
People’s Daily	人民日报	Ministerial Level (部级)	150Mil	151676
Xinhua News	新华社	Ministerial Level (部级)	108Mil	157363
QiuShi	求是	Vice Ministerial Level (副部级)	5.15Mil	36913
PLA Daily	解放军报	Vice Ministerial Level (副部级)	22.03Mil	38288
Guangming Daily	光明日报	Vice Ministerial Level (副部级)	25.28Mil	88959
Economic Daily	经济日报	Vice Ministerial Level (副部级)	5.75Mil	119277
China Daily	中国日报	Vice Ministerial Level (副部级)	64.98Mil	170154
Science and Technology Daily	科技日报	Vice Ministerial Level (副部级)	1.13Mil	46842
China Global Television Network	中国国际电视台	TODO: find out about this	6.01Mil	81839
China Central Television	中央电视台	Vice Ministerial Level (副部级)	131Mil	162092
China National Radio	中央人民广播电台	Vice Ministerial Level (副部级)	5.30Mil	104204
Beijing Daily	北京日报	TODO: find out about this	10.67Mil	114386
Workers’ Daily	工人日报	Bureau Level (局级)	3.79Mil	53219
China Youth Daily	中国青年报	Bureau Level (局级)	20.20Mil	126752
China Women’s News	中国妇女报	Bureau Level (局级)	11.76Mil	78255
Legal Daily	法治日报	Bureau Level (局级)	4.98Mil	94596
China News Service	中国新闻社	Bureau Level (局级)	17.59Mil	82256
Farmers’ Daily	农民日报	Bureau Level (局级)	0.435Mil	35704

Figure 11: Summary of State Run Media

C GDELT Data

D Additional Regression Results

	Mean	SD	Min	Median	Max	N
Likes	8747.01	59005.04	0.00	506.00	6.0e+06	2.6e+05
Comments	859.02	4134.36	0.00	138.00	5.5e+05	2.6e+05
Shares	3599.49	1.5e+05	0.00	197.00	4.0e+07	2.6e+05

Table 20: Weibo Audience Attention and Engagement Summary Statistics

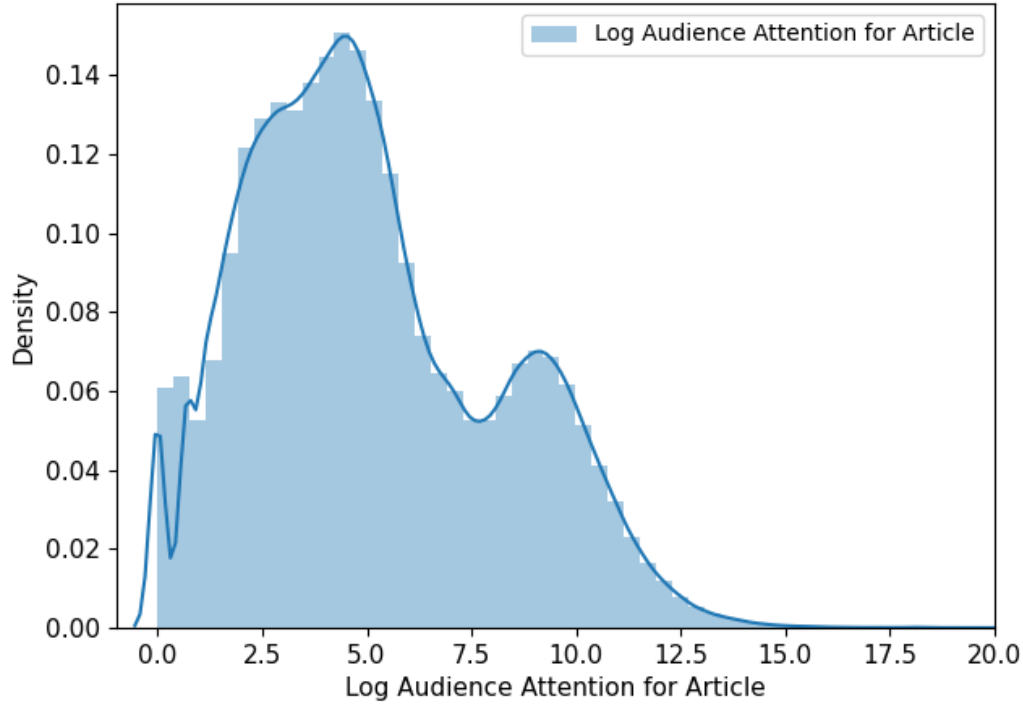


Figure 12: Log Audience Attention towards Weibo Post, As Measured by Likes, Comments, and Shares

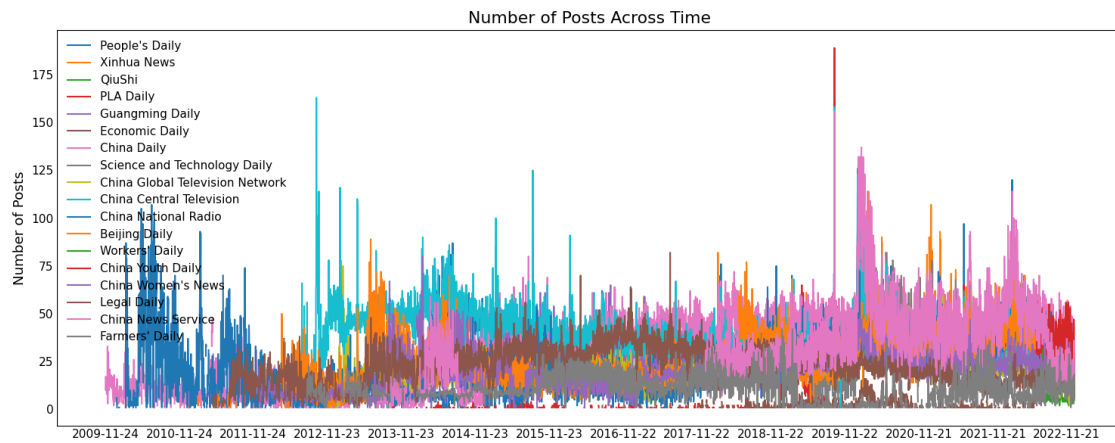


Figure 13: Number of Posts per Day Across Time

Provinces	Counts
Beijing	33896
Gansu	31765
Guangdong	4639
Jiangxi	3986
Shanghai	3825
Hubei	3106
Zhejiang	2532
Jiangsu	2081
Fujian	1979
Jilin	1854
Liaoning	1676
Anhui	1647
Xizang	1627
Sichuan	1300
Henan	987
Tianjin	875
Shaanxi	848
Hainan	786
Yunnan	668
Guizhou	640
Hunan	599
Chongqing	477
Guangxi	463
Shandong	429
Nei Mongol	347
Heilongjiang	300
Hebei	291
Shanxi	227
Xinjiang Uygur	166
Ningxia Hui	95
Qinghai	79

Table 21: Geographical Distribution of Verbal and Material Conflicts

Event Class	Count	Percent
Verbal Conflict	55539	45.54%
Material Conflict	48651	39.89%
Material Cooperation	17772	14.57%
Verbal Cooperation	0	0%

Table 22: Event Distribution by Verbal vs. Material and Cooperation vs. Conflict

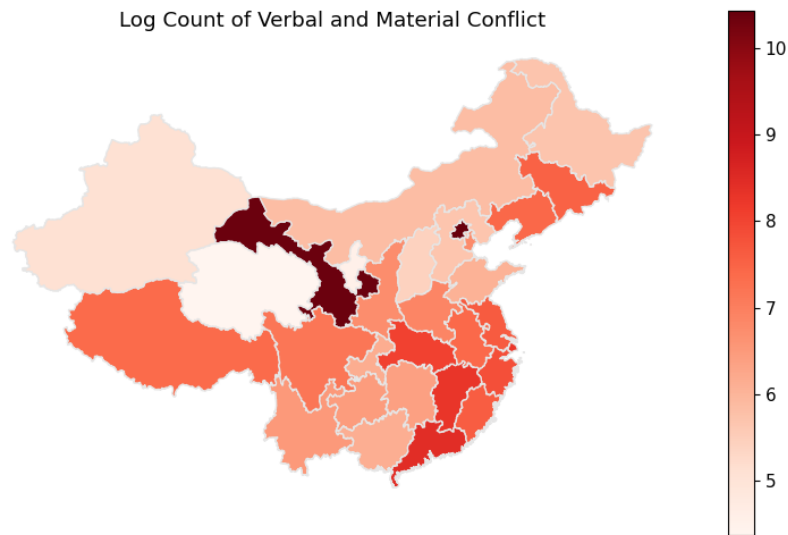


Figure 14: GDELT Events Geographical Heatmap

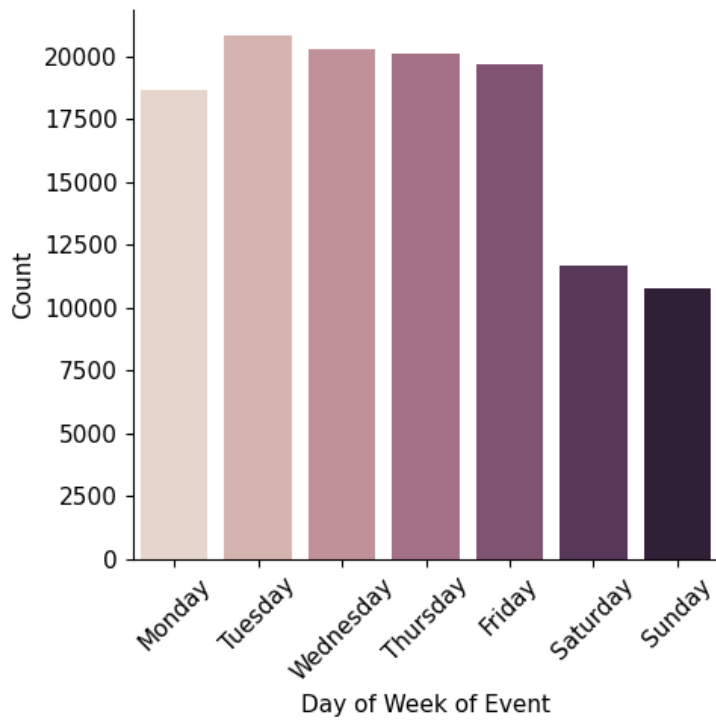


Figure 15: Day of Week of Event Occurrence

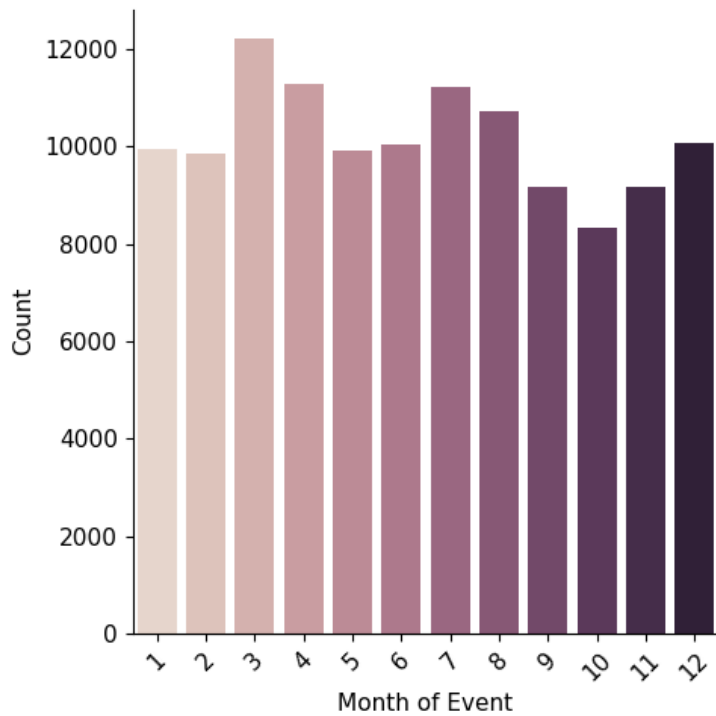


Figure 16: Month of Event Occurrence

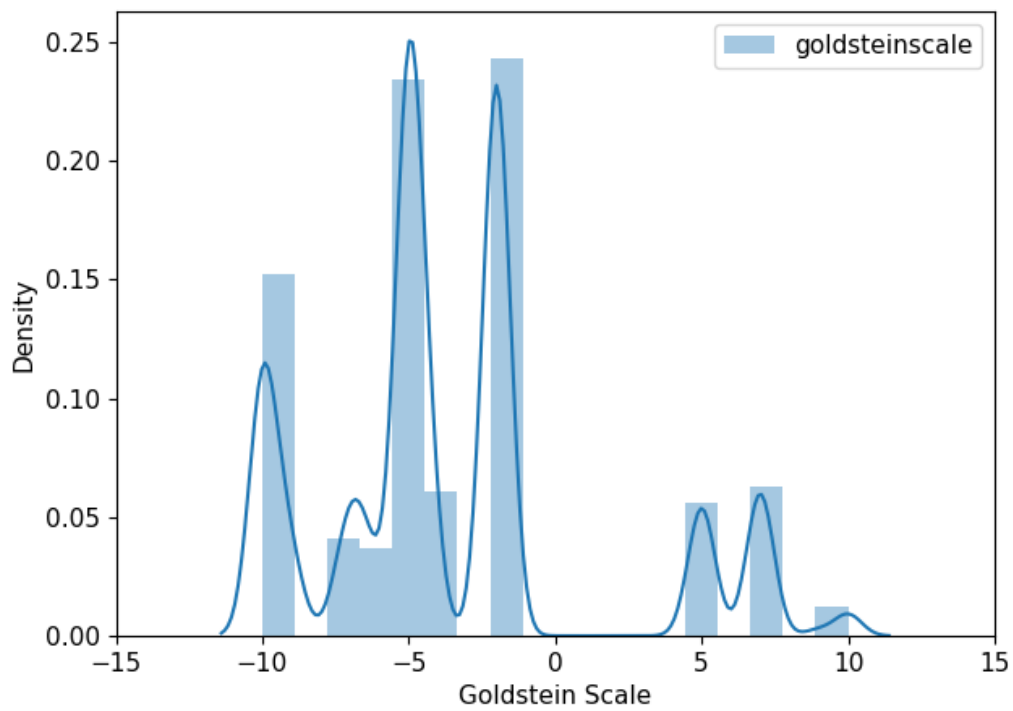


Figure 17: Distribution of Goldstein Scores of GDELT Events

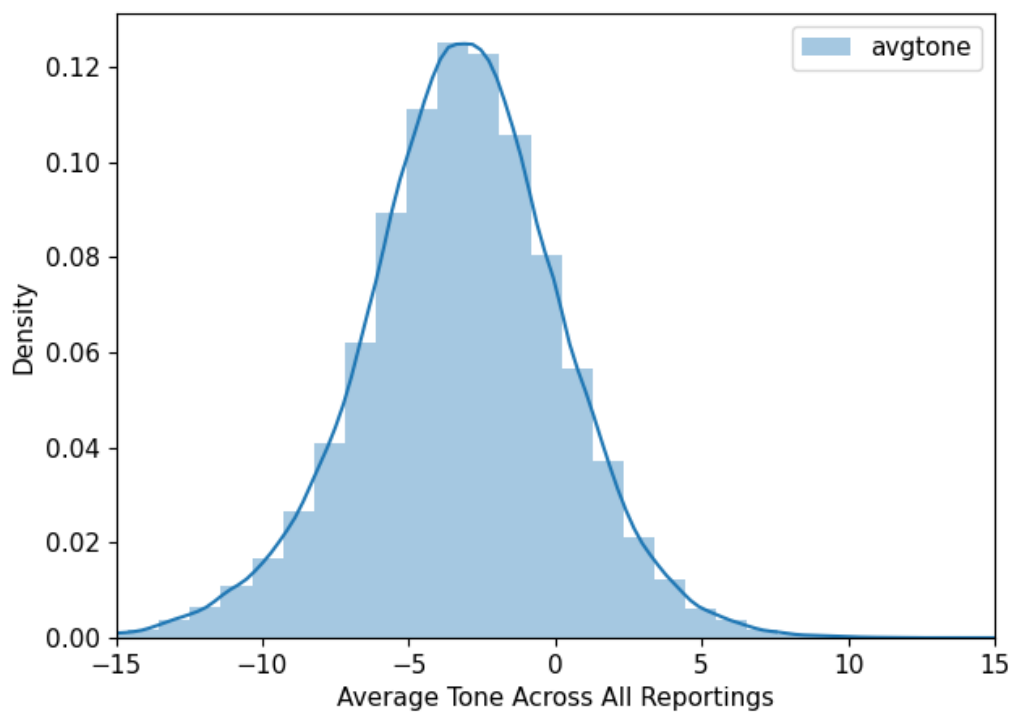


Figure 18: Distribution of Average Tone of GDELT Events

	Mean	SD	Min	Median	Max	N
goldsteinscale	-3.45	4.84	-10.00	-5.00	10.00	1.2e+05
avgtone	-3.22	3.47	-35.71	-3.15	17.13	1.2e+05
numarticles	4.07	3.99	1.00	3.00	330.00	1.2e+05

Table 23: Summary Statistics for Continuous Variables

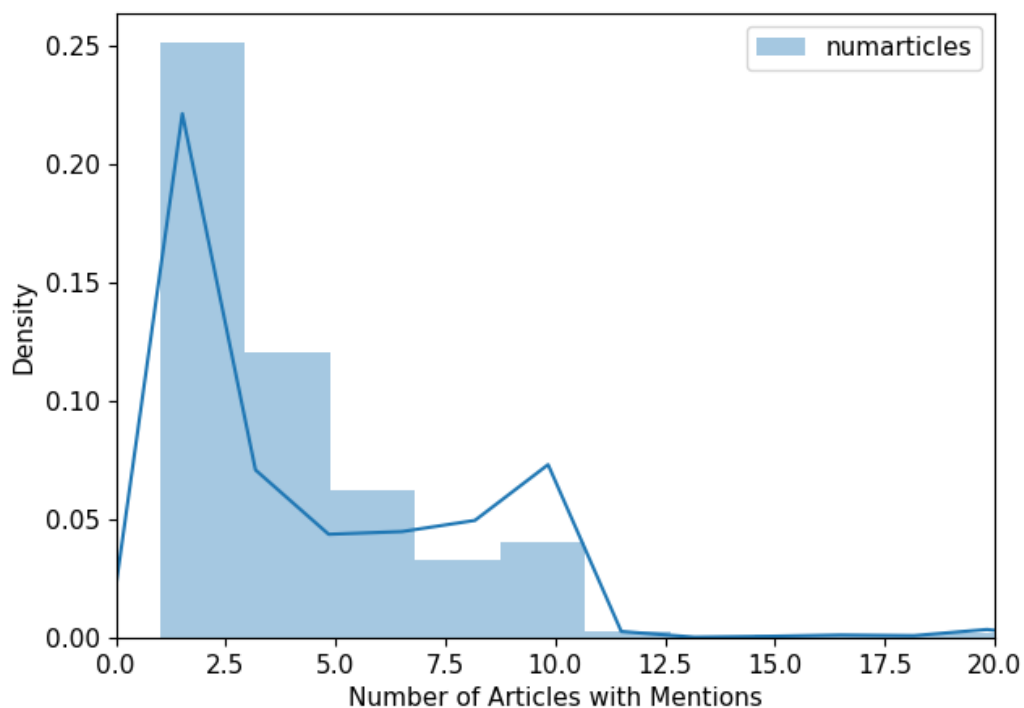


Figure 19: Distribution of Number Article Mentions of GDELT Events

	Quad Class		Goldstein Score	Average Tone
	Counts	Percent	Mean (SD)	Mean (SD)
2	17772	14.57175	6.40 (1.44)	-1.80 (3.28)
3	55539	45.53795	-3.33 (1.72)	-3.04 (3.31)
4	48651	39.89029	-7.17 (2.29)	-3.94 (3.53)
Total	121962	100	-3.45 (4.84)	-3.22 (3.47)

Table 24: Goldstein Scale and Average Tone by Quadclass

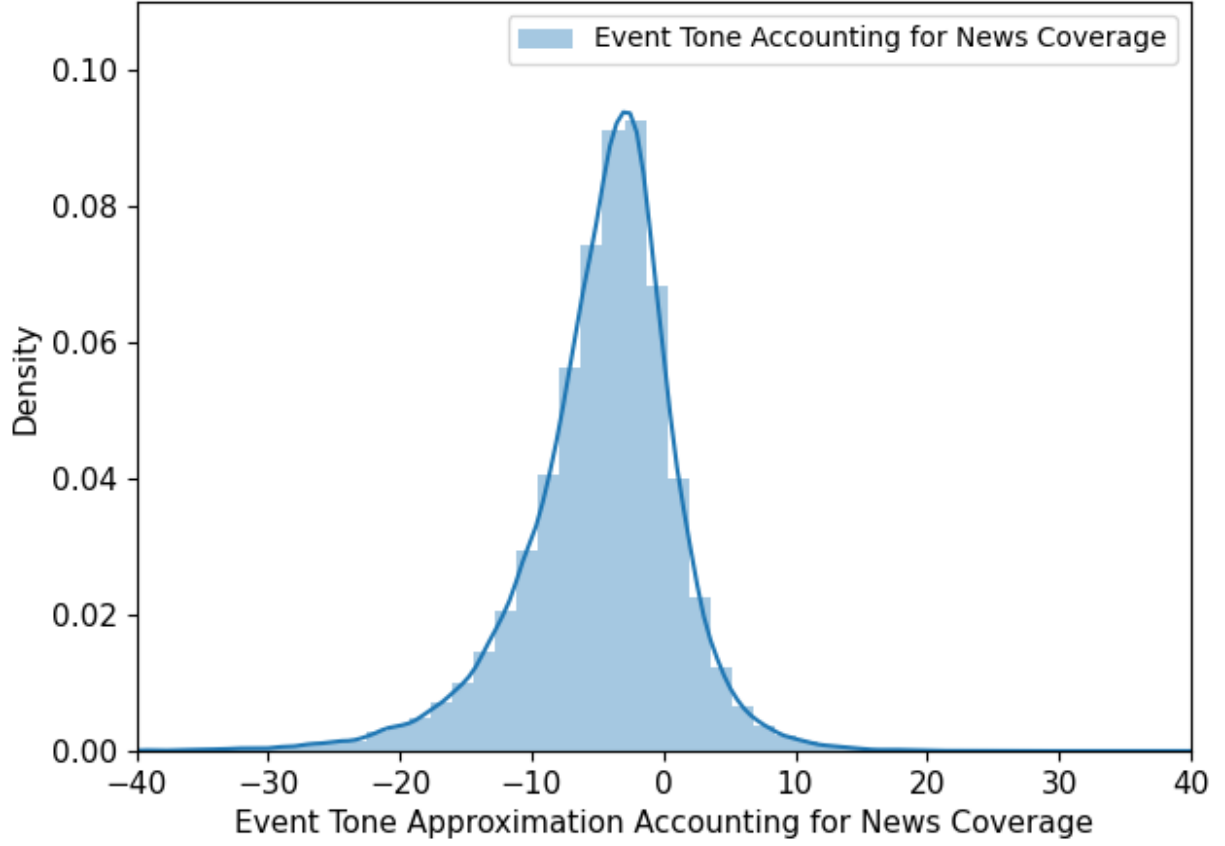


Figure 20: Tone Approximation Accounting for News Coverage for GDELT Events

	All		Quadclass = 2		Quadclass = 3		Quadclass = 4	
	Counts	Percent	Counts	Percent	Counts	Percent	Counts	Percent
Goldstein Score	-3.45	4.84	6.40	1.44	-3.33	1.72	-7.17	2.29
Average Tone	-3.22	3.47	-1.80	3.28	-3.04	3.31	-3.94	3.53
numarticles	4.07	3.99	3.97	3.28	4.00	3.98	4.19	4.22

Table 25: GDELT Continuous Variables by Quadclass

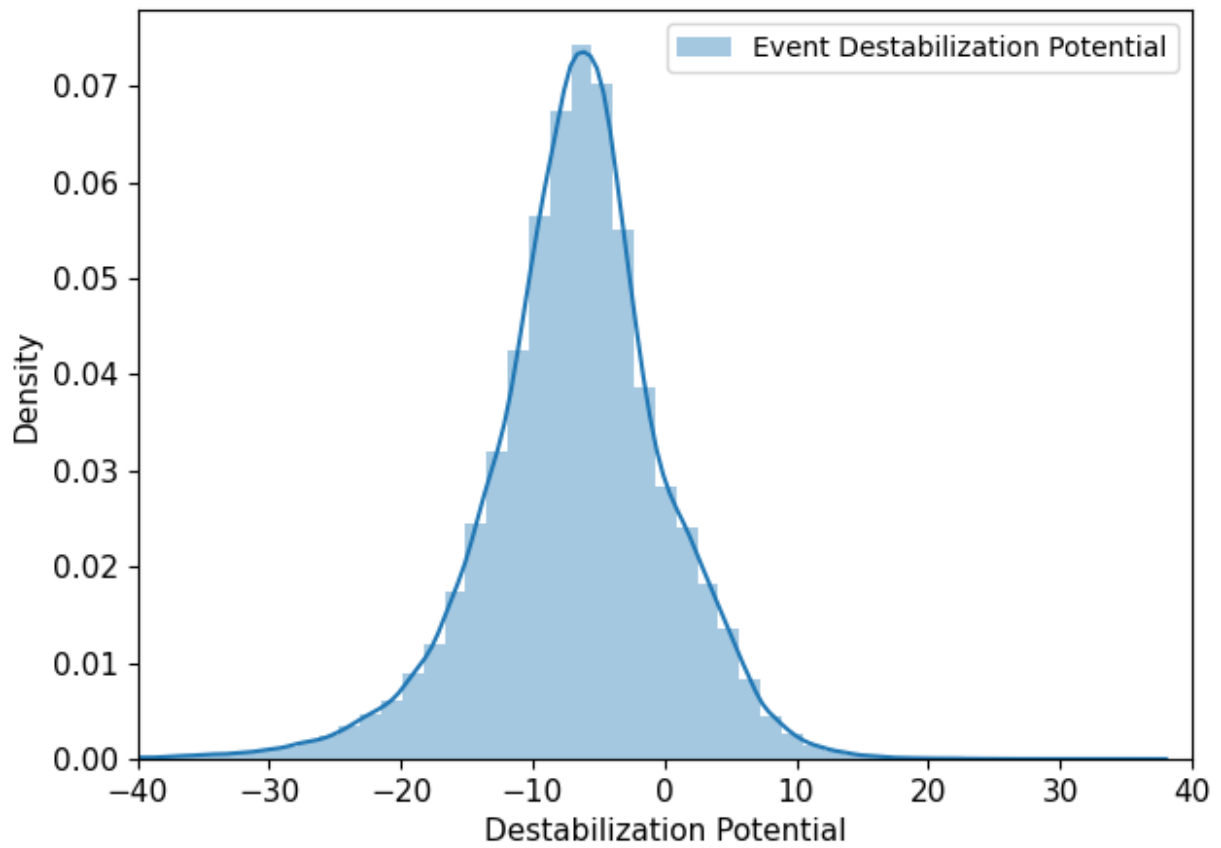


Figure 21: Destabilization Potential for GDELT Events

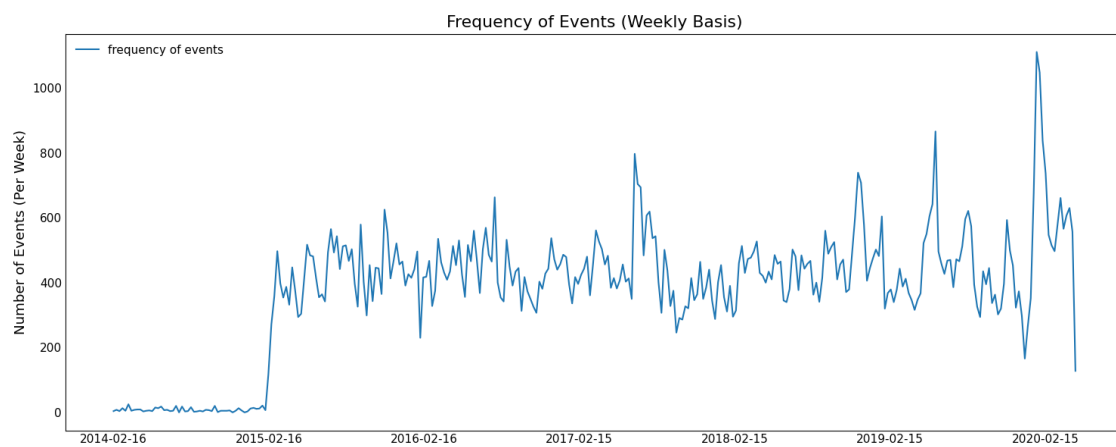


Figure 22: GDELT Events per Week

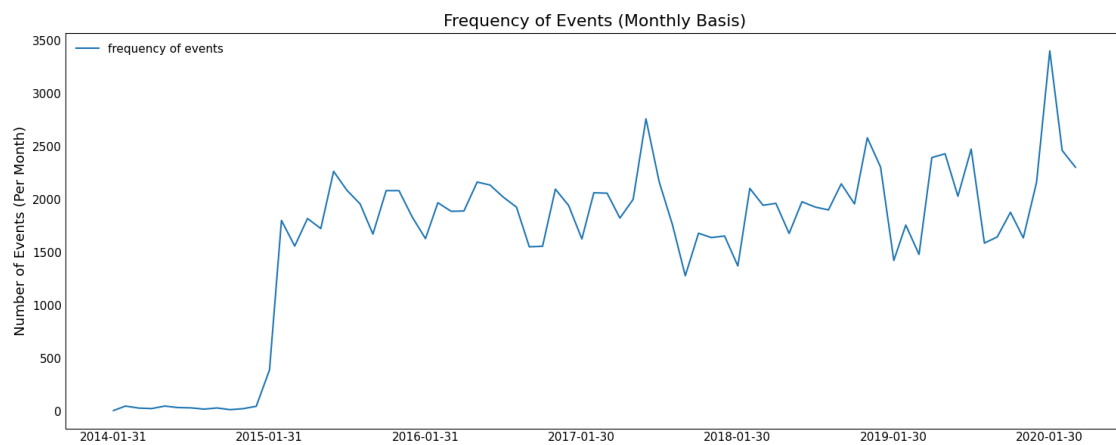


Figure 23: GDELT Events Per Month

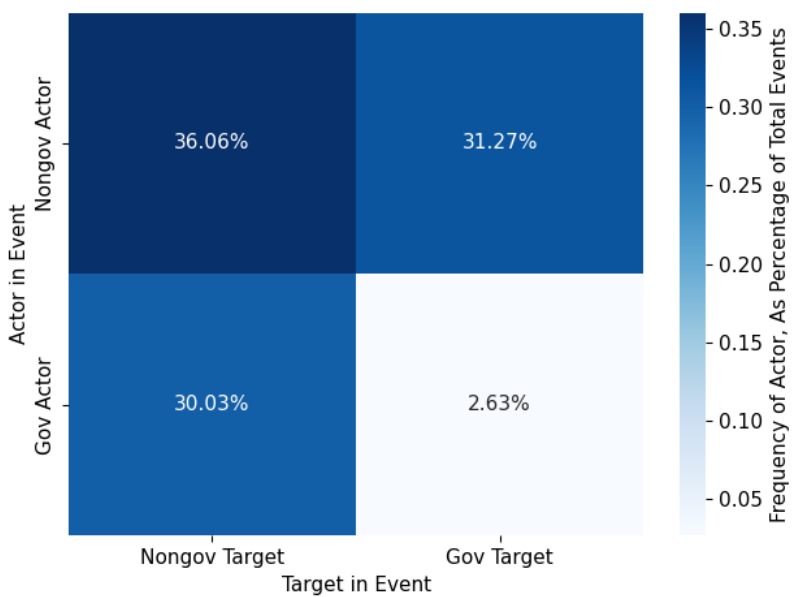


Figure 24: Unrest Event Government vs. Non-Government Actor and Target

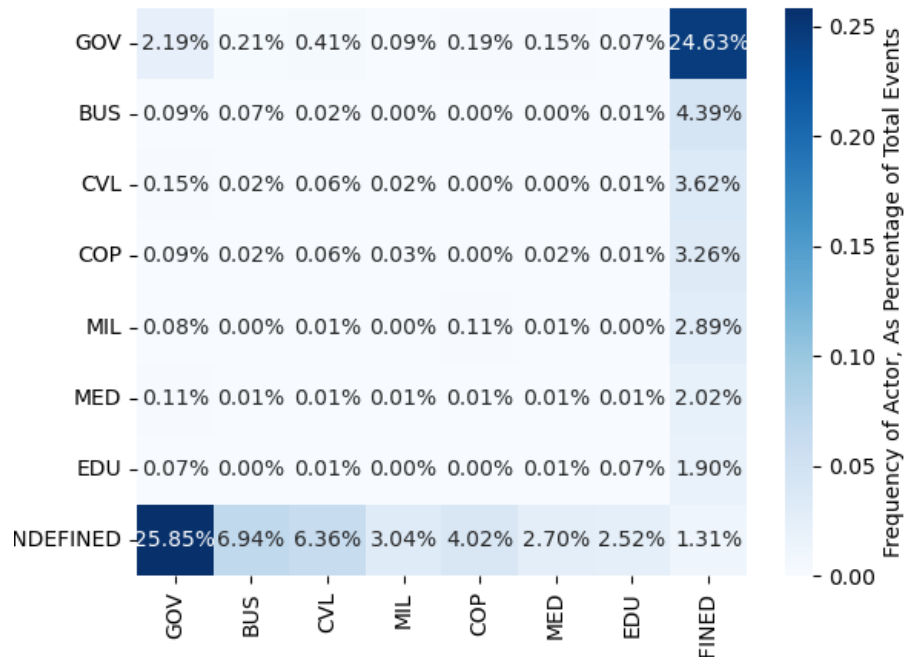


Figure 25: Unrest Event Actor and Target Type (probabilities calculated by actor)

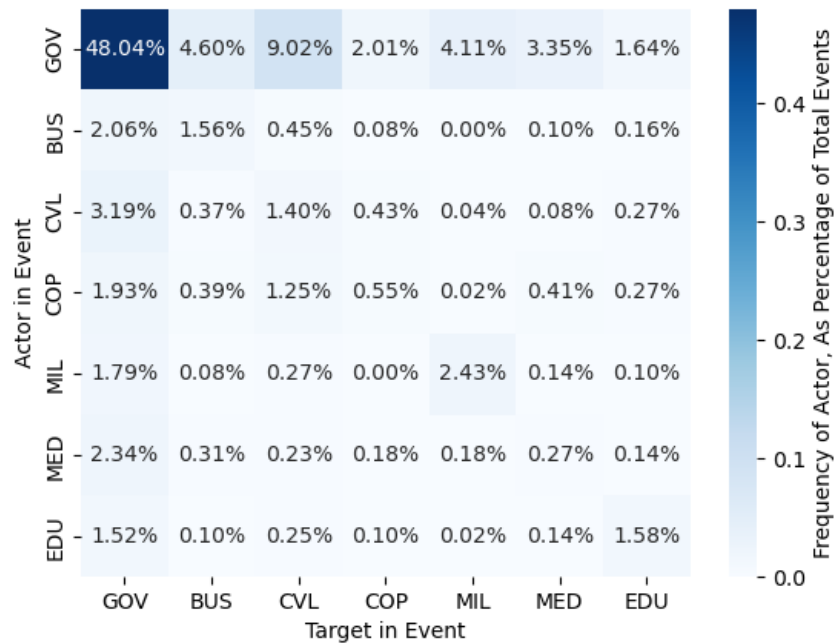


Figure 26: Unrest Event Actor and Target Type (probabilities calculated by target)

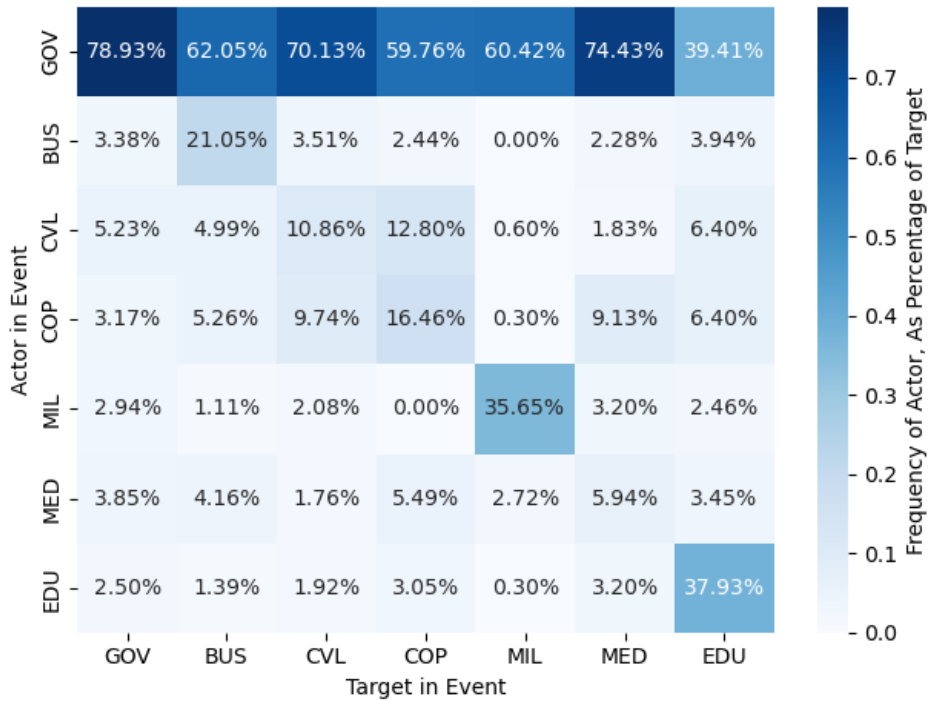


Figure 27: Unrest Event Actor and Target Type (probabilities calculated by target)

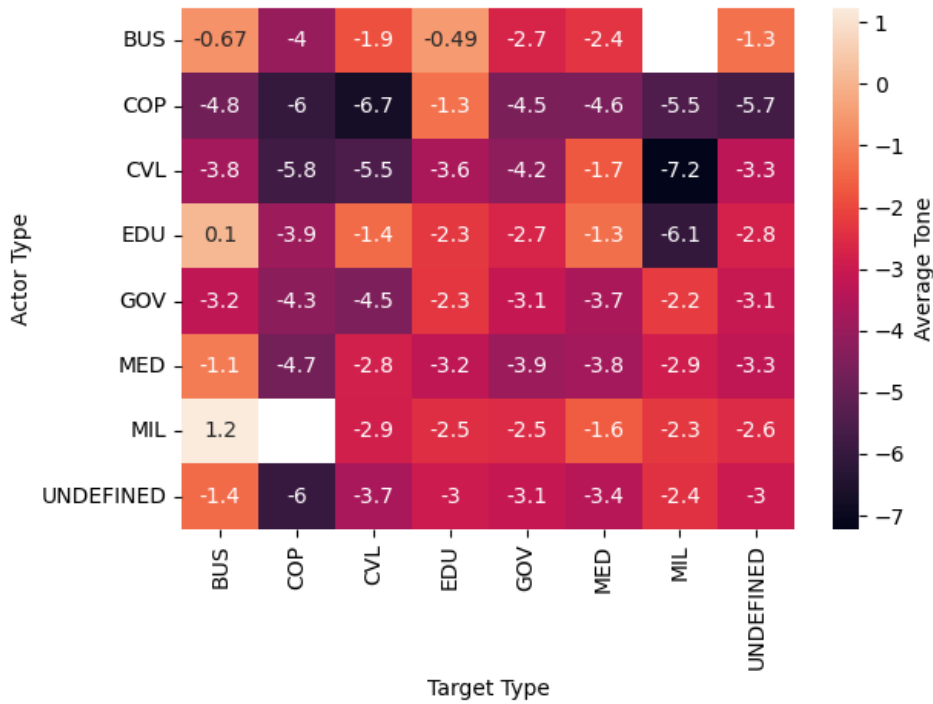


Figure 28: Average Tone by Actor and Target Type

	Actor of Event		Target of Event	
	Counts	Percent	Counts	Percent
AGR	122	0.21	155	0.23
BUS	4958	8.56	7841	11.39
COP	3794	6.55	3454	5.02
CRM	289	0.50	456	0.66
CVL	4169	7.20	7461	10.84
EDU	2229	3.85	2928	4.25
ELI	689	1.19	683	0.99
ENV	10	0.02	10	0.01
GOV	30248	52.22	30865	44.84
HLH	650	1.12	626	0.91
HRI	70	0.12	87	0.13
IGO	2	0.00	2	0.00
IMG	2	0.00		
JUD	1506	2.60	1517	2.20
LAB	512	0.88	796	1.16
LEG	820	1.42	602	0.87
MED	2361	4.08	3122	4.54
MIL	3348	5.78	4650	6.76
MNC	730	1.26	1162	1.69
NGO	78	0.13	74	0.11
OPP	864	1.49	1535	2.23
RAD	8	0.01	24	0.03
REB	97	0.17	99	0.14
REF	39	0.07	51	0.07
SEP	58	0.10	157	0.23
SET	5	0.01	8	0.01
SPY	122	0.21	261	0.38
UAF	142	0.25	204	0.30
Total	57922	100.00	68830	100.00

Table 26: GDELT Summary Statistics - Actor Type

	(1)	(2)	(3)	(4)	(5)	(6)
Mean Goldstein Scale	-9.90 (13.99)	32.74** (8.21)	101.91* (47.53)	5.44 (14.49)	-376.81 (2057.14)	28.23** (9.59)
Mean Average Tone	34.11* (16.99)	-11.19 (9.66)	-84.68+ (49.00)	30.90 (21.45)	439.44 (2292.02)	4.10 (11.99)
Day of Week	-41.66** (8.50)	-42.45** (7.13)	-43.73** (16.08)	-67.30** (18.02)	550.10 (2979.23)	-70.69** (14.54)
Days Linear	1.42** (0.26)	0.82** (0.06)	-0.16 (0.43)	1.05** (0.19)	9.72 (46.74)	0.82** (0.07)
Month	-15.10** (4.75)	-33.63** (5.19)	-63.70* (26.22)	-33.24** (8.13)	-0.97 (121.97)	-17.62** (4.59)
Fraction Negative Given Domestic Political	7.89** (2.43)					
Fraction Neutral Given Domestic Political		-4.88** (1.25)				
Fraction Positive Given Domestic Political			12.80+ (7.23)			
Fraction Domestic Given Political				-15.84* (6.43)		
Fraction Negative Domestic Political					517.27 (2649.98)	
Fraction Political						-6.30** (1.94)
Constant	-1324.14 (811.39)	3361.54** (526.61)	-1837.38 (1777.48)	15105.00** (5596.74)	-37357.91 (198082.14)	2954.52** (510.05)
Obs	2260	2260	2260	2260	2260	2260
Adj. R ²	-0.58	-0.09	-4.32	-1.74	-436.90	-0.58
F-stat	635.51	922.77	188.89	365.91	2.29	635.01

Table 27: IV2SLS Reduced Form: Log Aggregate Attention Potential

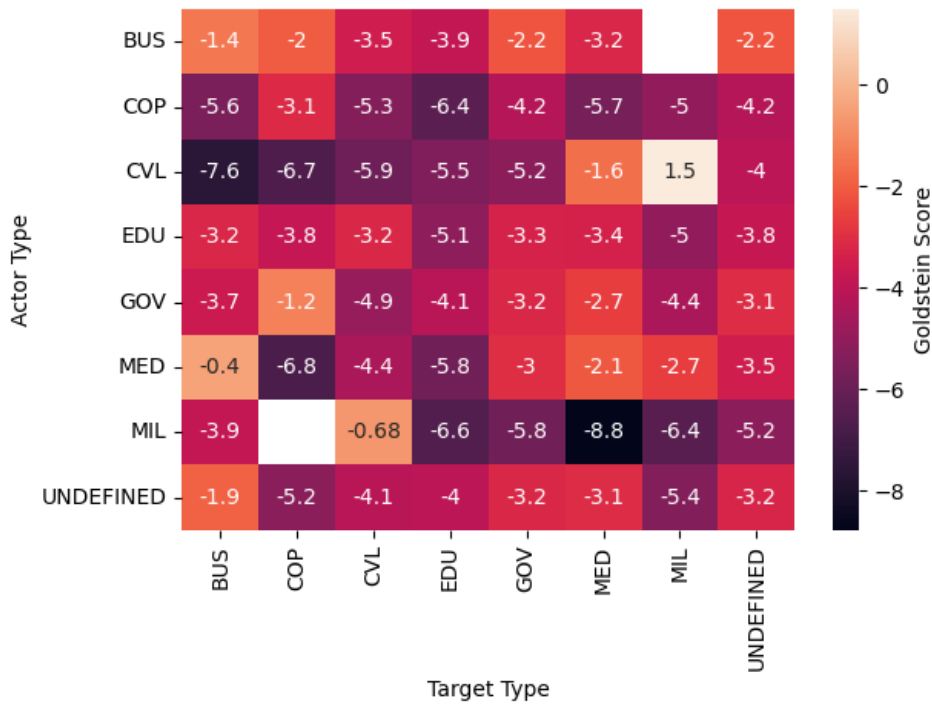


Figure 29: Goldstein Score by Actor and Target Type

References

- Akhmedov, Akhmed, and Ekaterina Zhuravskaya.** 2004. “Opportunistic Political Cycles: Test in a Young Democracy Setting*.” *The Quarterly Journal of Economics*, 119(4): 1301–1338.
- Alrababa’h, Ala’ , and Lisa Blaydes.** 2021. “Authoritarian media and diversionary threats: lessons from 30 years of Syrian state discourse.” *Political Science Research and Methods*, 9(4): 693–708.
- Barbera, Pablo, Anita R. Gohdes, Evgeniia Iakhnis, and Thomas Zeitzof.** 2020. “Distract and divert: How world leaders use social media during contentious politics.”
- Beraja, Martin, Andrew Kao, David Y Yang, and Noam Yuchtman.** 2021. “AI-tocracy.” National Bureau of Economic Research Working Paper 29466.
- Bernard, Victor L., and Jacob K. Thomas.** 1989. “POST-EARNINGS-ANNOUNCEMENT DRIFT - DELAYED PRICE RESPONSE OR RISK PREMIUM.” *Journal of Accounting Research*, 27: 1–36.

- Besley, Timothy, and Andrea Prat.** 2006. “Handcuffs for the Grabbing Hand? Media Capture and Government Accountability.” *American Economic Review*, 96(3): 720–736.
- Besley, Timothy, and Sacha Dray.** 2023. “The political economy of lockdown: Does free media matter?” *European Journal of Political Economy*, 102361.
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer.** 2022. “Salience.” *Annual Review of Economics*, 14(1): 521–544.
- Brace, Paul, and Barbara Hinckley.** 1993. “Presidential Activities from Truman through Reagan: Timing and Impact.” *The Journal of Politics*, 55(2): 382–398.
- Carter, Erin Baggott.** 2020. “Diversionary cheap talk: economic conditions and US foreign policy rhetoric, 1945-2010.” *International Interactions*, 46(2): 163–198.
- Chetty, Raj, Adam Looney, and Kory Kroft.** 2009. “Salience and Taxation: Theory and Evidence.” *American Economic Review*, 99(4): 1145–77.
- Cohen, Lauren, and Andrea Frazzini.** 2008. “Economic Links and Predictable Returns.” *The Journal of Finance*, 63(4): 1977–2011.
- Dellavigna, Stefano, and Joshua M. Pollet.** 2009. “Investor Inattention and Friday Earnings Announcements.” *The Journal of Finance*, 64(2): 709–749.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova.** 2018. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” *CoRR*, abs/1810.04805.
- Djankov, Simeon, Caralee McLiesh, Tatiana Nenova, and Andrei Shleifer.** 2003. “Who Owns the Media?” *Journal of Law and Economics*, 46(2): 341–81.
- Durante, Ruben, and Ekaterina Zhuravskaya.** 2018. “Attack When the World Is Not Watching? US News and the Israeli-Palestinian Conflict.” *Journal of Political Economy*, 126(3): 1085–1133.

- Egorov, Georgy, Sergei Guriev, and Konstantin Sonin.** 2009. “Why Resource-poor Dictators Allow Freer Media: A Theory and Evidence from Panel Data.” *The American Political Science Review*, 103(4): 645–668.
- Eisensee, Thomas, and David Strömberg.** 2007. “News Droughts, News Floods, and U. S. Disaster Relief*.” *The Quarterly Journal of Economics*, 122(2): 693–728.
- Gelpi, Christopher.** 1997. “Democratic Diversions: Governmental Structure and the Externalization of Domestic Conflict.” *Journal of Conflict Resolution*, 41(2): 255–282.
- Gentzkow, Matthew, Edward L Glaeser, and Claudia Goldin.** 2004. “The Rise of the Fourth Estate: How Newspapers Became Informative and Why It Mattered.” National Bureau of Economic Research Working Paper 10791.
- Gohdes, Anita R.** 2020. “Repression Technology: Internet Accessibility and State Violence.” *American Journal of Political Science*, 64(3): 488–503.
- Goncharov, Maxim.** n.d.. “The Dark Side of Social Media.”
- Hensel, Lukas, Marc Witte, Stefano Caria, Thiemo Fetzer, Stefano Fiorin, Friedrich M Götz, Margarita Gómez, Johannes Haushofer, Andriy Ivchenko, Gordon Kraft-Todd, and et al.** 2020. “Global Behaviors, Perceptions, and the Emergence of Social Norms at the Onset of the COVID-19 Pandemic.”
- Hobbs, William R., and Margaret E. Roberts.** 2018. “How Sudden Censorship Can Increase Access to Information.” *American Political Science Review*, 112(3): 621–636.
- Holland, Alisha C.** 2016. “Forbearance.” *American Political Science Review*, 110(2): 232–246.
- James M. Snyder, Jr., and David Strömberg.** 2008. “Press Coverage and Political Accountability.” National Bureau of Economic Research, Inc NBER Working Papers 13878.
- James, Patrick.** 1990. “Crisis and War.” *American Political Science Review*, 84(1): 362–363.
- Johnson, Rie, and Tong Zhang.** 2017. “Deep Pyramid Convolutional Neural Networks for Text Categorization.” 562–570. Vancouver, Canada: Association for Computational Linguistics.

- Joulin, Armand, Edouard Grave, Piotr Bojanowski, and Tomáš Mikolov.** 2016. “Bag of Tricks for Efficient Text Classification.” *CoRR*, abs/1607.01759.
- Kanat, Kilic.** 2014. “Diversionary Foreign Policy in Authoritarian States: The Use of Multiple Diversionary Strategies by Saddam Hussein during the Gulf War.” *Journal of Strategic Security*, 7: 16–32.
- Kim, Yoon.** 2014. “Convolutional Neural Networks for Sentence Classification.” *CoRR*, abs/1408.5882.
- King, Gary, Jennifer Pan, and Margaret E. Roberts.** 2013. “How Censorship in China Allows Government Criticism but Silences Collective Expression.” *American Political Science Review*, 107(2 (May)): 1–18. Please see our followup article published in Science, “Reverse-Engineering Censorship In China: Randomized Experimentation And Participant Observation.”
- 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.
- Lai, Siwei, Liheng Xu, Kang Liu, and Jun Zhao.** 2015. “Recurrent Convolutional Neural Networks for Text Classification.” *Proceedings of the AAAI Conference on Artificial Intelligence*, 29(1).
- Liu, Pengfei, Xipeng Qiu, and Xuanjing Huang.** 2016. “Recurrent Neural Network for Text Classification with Multi-Task Learning.” *CoRR*, abs/1605.05101.
- Marra, Robin F., Charles W. Ostrom, and Dennis M. Simon.** 1990. “Foreign Policy and Presidential Popularity: Creating Windows of Opportunity in the Perpetual Election.” *The Journal of Conflict Resolution*, 34(4): 588–623.
- Martin, Hans-Peter.** 1997. *The global trap: globalization and the assault on prosperity and democracy*. HSRC/RGN; Pluto Press; White Lotus; Zed Books Pretoria: Annandale, N.S.W.: Bangkok: London; New York.

- Morgan, T. Clifton, and Christopher J. Anderson.** 1999. “Domestic Support and Diversionary External Conflict in Great Britain, 1950- 1992.” *The Journal of Politics*, 61(3): 799–814.
- Morgan, T. Clifton, and Kenneth N. Bickers.** 1992. “Domestic Discontent and the External Use of Force.” *Journal of Conflict Resolution*, 36(1): 25–52.
- Nordhaus, William D.** 1975. “The Political Business Cycle.” *The Review of Economic Studies*, 42(2): 169–190.
- Oakes, Amy.** 2006. “Diversionary War and Argentina’s Invasion of the Falkland Islands.” *Security Studies*, 15(3): 431–463.
- Roberts, Margaret E.** 2018. *Censored: Distraction and diversion inside China’s Great Firewall*. Princeton University Press.
- Sims, Christopher A.** 2003. “Implications of rational inattention.” *Journal of Monetary Economics*, 50(3): 665–690.
- Sobek, David.** 2007. “Rallying Around the Podesta: Testing Diversionary Theory Across Time.” *Journal of Peace Research*, 44(1): 29–45.
- Sprecher, Christopher, and Jr. Karl Derouen.** 2002. “Israeli Military Actions and Internalization-externalization Processes.” *Journal of Conflict Resolution*, 46(2): 244–259.
- Sun, Yu, Shuohuan Wang, Yu-Kun Li, Shikun Feng, Xuyi Chen, Han Zhang, Xin Tian, Danxiang Zhu, Hao Tian, and Hua Wu.** 2019. “ERNIE: Enhanced Representation through Knowledge Integration.” *CoRR*, abs/1904.09223.
- Swinhoe, Dan.** 2022. “Kazakhstan government shuts down Internet amid protests.” *Datacenter Dynamics*.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin.** 2017. “Attention Is All You Need.” *CoRR*, abs/1706.03762.
- Wen, Jaya.** 2022. “State Employment as a Strategy of Autocratic Control in China.”

Zhou, Peng, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. 2016. “Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification.” 207–212. Berlin, Germany: Association for Computational Linguistics.

Zuboff, Shoshana. 2018. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* . 1st ed.