Who Not What:

The Logic of China's Information Control Strategy[*]

Mary Gallagher[†]   Blake Miller[‡]

April 22, 2021

Abstract

In this paper, we examine how the Chinese state controls social media. While social media companies are responsible for censoring their platforms, they also selectively report certain users to the government. Our focus in this article is to understand the logic behind media platforms' decisions to report users or content to the government. We find that content is less relevant than commonly thought. Information control efforts often focus on *who* is posting, rather than *what* they are posting. The state permits open discussion and debate on social media while controlling and managing influential social forces that may challenge the Party-State's hegemonic position. We build on "ideology and organization" by Schurmann (1968), emphasizing the Party's goals of embedding itself in all social structures and limiting the ability of non-Party individuals, networks, or groups to carve out separate space for leadership and social status. In the virtual public sphere, the Chinese Communist Party (CCP) continues to apply these principles to co-opt, repress, and limit the reach of influential non-Party "thought leaders." We find evidence to support this logic through qualitative and quantitative analysis of leaked censorship documents from a social media company and government documents on information control.

摘要：

此论文探讨中国政府如何控制社交媒体。社交媒体公司不仅负责审查自身平台，也会有选择性地向政府部门上报一些用户。此论文重点研究媒体平台上报用户或内容背后的选择逻辑。我们发现媒体对内容的审查并非如过往研究发现的那么重要。信息审查通常更着眼于发布者是谁，而不是发布者说了什么。中国政府允许社交媒体上的公开讨论和辩论，而同时控制和管理可能挑战党国领导权的意见领袖和

[†]Professor, Department of Political Science, University of Michigan, Haven Hall, Ann Arbor, MI 48109-1045 (E-mail: metg@umich.edu).

[‡]Corresponding author, Department of Methodology, London School of Economics and Political Science, Columbia House, Aldwych, London WC2A 2AE (E-mail: b.a.miller@lse.ac.uk).

有影响力的社会势力。我们的研究基于舒尔曼的《共产主义中国的意识形态和组织》。舒尔曼强调党的目标是将自身融入所有社会结构，并限制其他个人或团体能占领的公共空间。中国共产党根据这些原则来吸纳，压制和限制体制外"意见领袖"在网络公共空间里的影响力。我们使用定性与定量分析研究一整套被泄露的新浪微博审查文件和一整套思想工作干部指南，分析结果证实了我们的理论。

关键词： 社交媒体; 意见领袖; 压制; 审查

Introduction

Neihan Duanzi (neihan duanzi, 内涵段子) was a seemingly harmless social media portal for videos, memes, and jokes in China, similar to Reddit or 4chan. It rarely hosted criticisms of the government or calls for collective action. However, as the site gained popularity, users began to form offline communities, meeting up and identifying each other in public, sometimes communicating in secret codes of car honks at intersections. In April of 2018, the government surprised many by shutting the site down. We argue that this decision reflects the regime's strategy of maximally shrinking non-official public spaces that could threaten the CCP's authority or compete with it ideologically in the future. Neihan Duanzi's large and growing community was not a clear threat to the CCP at the time of its ban, but it had the potential to transform into one. Using the example of another platform, Sina Weibo (xinlang weibo, 新浪微博), we argue that the key to shrinking this space is the careful targeting of individuals who are influential within counter-hegemonic spaces such as Neihan Duanzi.

While the state could reassert control over these spaces by shuttering all private social media platforms, it benefits from a well-maintained and constrained private media ecosystem. The Chinese state, once at odds with capitalism, is now using it as a source of stability. Drawing on the expertise of tech firms who continue to innovate in social networks, digital payments, and targeted advertising analytics, the government can reserve real-world and violent repression for those who are flagged by algorithms predicting future acts of state subversion. These systems are already in development (Meissner, 2017; Yang, Yang, & Fei, 2017). Through co-optation of technology companies, the state enforces information control and repression with a scalpel rather than a hammer.

The Chinese state has carefully constructed its repression regime to tame citizens via impersonal censorship algorithms, calibrating, and customizing its tactics for different individuals and contexts. For many, the intrusion of state security into their lives is merely virtual and often covert and unobserved. For a select few who dare to push the boundaries, online censorship may be more explicit and threatening. It may also move from the virtual realm to real-world interaction—home visits, invitations to "drink tea," interrogations, detentions in a black jail, or worse.

We analyze when delegated, sometimes automated, censorship of online content shifts back to the state's repressive apparatus. The Chinese government delegates online censorship to private internet content producers (ICPs) like Sina and Tencent, and compels them to satisfy the government's demand for information control. While much of the work of delegated censorship is the deletion or hiding of sensitive content, ICPs also selectively report certain content and users back to the government. We seek to understand the logic of this reporting system: who and what are targeted for handling by the state itself?

We find that "reporting up"—which likely results in direct state repression or co-optation—is targeted toward influential public opinion leaders whose standing and influence may threaten the Party's hegemonic presence in China's online public sphere. Substantive topics, though relevant, matter less than user influence and virality in state efforts to "guide opinion" through direct repressive interventions. Many topics—including topics that have an apparent collective action threat (a political dissident, a workers' strike)—are not reported up unless they achieve a certain degree of virality or influence. By contrast, many viral discussions that challenge the legitimacy of the regime in more indirect and oblique ways (elite defection, ridicule of top leaders) are reported up.

Work by Schurmann (1968) speaks to the state logic of "who not what." In his work on the early years of CCP rule, he emphasizes the importance that the CCP placed on ideology and organization in its social revolution, which he conceptualized as a process of destruction and replacement of former elites. Communist control had both ideational and spatial dimensions. Ideology replaced other values and norms of China's old society. Communist organization embedded the party in all social, economic, and political structures and eliminated the old elite, from landlords to factory owners. However, in the second edition of Ideology and Organization, Schurmann's analysis of the Maoist state during the Cultural Revolution is instructive for understanding how Maoism approaches challenges of a new, emerging elite. In Schurmann's analysis, social mobilization, especially of students in the Cultural Revolution, challenged the CCP's organizational dominance. "If power is the key element in organization, authority is the key element in a social system" (Schurmann, 1968, 521).

The contemporary context is quite different from that of the late 1960s. However, the emergence of a dynamic private sector, new social classes, and independent sources of wealth and social status poses similar challenges to the CCP's organizational dominance. In private firms, especially recently, the CCP has attempted to establish and strengthen the role of Party cells. In social media, however, organizational dominance is not possible. Instead, the CCP has developed tactics to reduce the authority of new voices, to limit their virality, and to eliminate or co-opt alternate sources of leadership and influence. The battle for online public opinion dictates an ambitious approach that seeks to allow for online expression while also suppressing the rise of influential voices and viral discussions. Although the government is worried about the Internet as a device to facilitate social movements and protests, it also worries simply when a large number of people online are talking about the same thing, what it calls "a public opinion emergency" (gonggong yulun weiji, 公共舆论危机) or an "internet opinion emergency" (wangluo yulun weiji, 网络舆论危机). Schurmann notes that in the context of an evolving social system, in which Communist organization is either not complete or weakened, authority is crucial. Controlling, managing, co-opting authoritative voices online is critical.

In social movement theory, authoritative actors can play the role of brokers in the diffusion of social mobilization, protest, and other forms of collective action. These influential actors serve as nodes in networks that may then connect up. In his study of transnational activism, Tarrow defines brokerage as "linking of two or more previously unconnected social actors by a unit that mediates their relations with one another and/or with yet other sites" (Tarrow, 2005, 190). In social media interaction, Big V-users can easily link up users who previously had no connection to each other, spawning information cascades and viral discussions of hot topics. As we show below, the government is especially worried about viral discussions sparked by influential thought leaders.

The state's evolving response to social grievances since the violent suppression of the student movement in 1989 can help us understand the tension between increased openness of communication and the state's goal of maintaining hegemony over discourse. Responding to the mistakes of 1989, the Chinese state developed a new toolkit—often called "responsive authoritarianism"—that combines targeted repression and broad responsiveness. It has become more tolerant of narrowly-expressed, localized, and economistic protests and grievances. At the same time, its ability to target and isolate political dissidents and

to use repression to silence the regime's most vociferous opponents has improved. With the explosion of social media since 2009, however, it has become more challenging for the state to prevent online publics from broadening beyond initially narrow grievance articulation. The networked and rapid nature of discussion online can unite similarly-aggrieved individuals across time, space, and barriers of region, class, and even language. For example, local scandals in food safety, wage arrears, or corruption can be more easily linked to systemic problems in the authoritarian political system, such as the lack of a free press, independent trade unions, or close relations between political and business elites. Social media has the potential to transform narrow, local problems to broad, system-challenging threats.

Maintaining responsiveness while containing the threat of the online public sphere is a delicate balancing act. To maintain its responsiveness, which seems to have vastly improved the government's legitimacy (Nathan, 2003; Dickson, 2016; Tang, 2016; Hasmath & Hsu, 2017), the central government uses social media discussion to monitor its local agents and mitigate the severe principal-agent problems inherent in China's top-down system. Ever concerned with information cascades that may lead to collective action (Kuran, 1991; Lohmann, 1994; DeNardo, 2014), the government seeks to prevent social media from acting as the "single spark that ignites a prairie fire" by controlling the virality influence of public opinion leaders online. CCP manuals to train Internet censors specifically reference this problem as the "butterfly effect" (hudie xiaoying, 蝴蝶效应), the need to control information before it gets out of (their) control (Wei et al., 2015; Gao & Zhang, 2011; Zou & Su, 2015; PRC State Council General Office, 2016).

This article has four parts. In the first section, we examine control dynamics over time, arguing that the regime has moved from broad-based repression to targeted repression coupled with responsiveness to a broader swathe of the population. The second section uses government manuals and documents on opinion guidance and internal censorship logs from a social media company to analyze when commercial censors report users and information back up to the government. These sources allow us to both quantitatively test our model and supplement our results with additional transcript evidence from government sources (Lorentzen, Fravel, & Paine, 2017). The third and final section discusses the findings.

1 The Rise of Responsive Authoritarianism and the Challenge of Virality

The Chinese Communist Party's approach to socio-political control has shifted over time. As economic reforms sped up in the 1990s, the government became far more tolerant toward protests that it saw as rooted in socio-economic grievances brought on by market reform. Small and narrow protests multiplied during this time, with protestors limiting their claims and often hiding organizations and leaders. These protests were different not only in who participated, but also in their claims, their targets, and their tactics. On the other side, the government began to overhaul the repressive apparatus of the state to improve its ability to preempt significant challenges while placating reasonable grievances (Tanner, 2004).

Protesters often framed their grievances as narrow and economistic. They focused on economic deprivation and loss precipitated by the disruptive and destabilizing economic restructuring and often unregulated growth. These protests overwhelmingly, and perhaps even intentionally, articulated their targets as bad local actors, such as corrupt local officials, greedy enterprise bosses, or compromised local bureaucrats who failed to enforce central laws strictly when such laws interfered with growth. These lower-level targets were best attacked using tactics that were wholeheartedly moralistic and normative, but also, at times, legalistic. As O'brien & Li (2006) have described, their invocation of resistance was rooted in support for the legitimate policies and laws of the central government and their anger that local governments did not enforce them. Demands were for better rule, not for a new set of rulers.

The turn toward increased state-society dialogue has been labeled bargained authoritarianism (Lee & Zhang, 2013), responsive authoritarianism (Heurlin, 2017), and contentious authoritarianism (Chen, 2012). These terms—though they emphasize different aspects of the phenomenon—hone in on the notion of protest as an accepted mode of political participation in contemporary China (Tang, 2016). State and society have both learned new tactics to make high levels of protest in an authoritarian regime sustainable, yielding neither to mass repression or a tipping point toward revolution (Heurlin, 2017; Lee & Zhang, 2013; Su & He, 2010; Cai, 2008). In this game of cat and mouse between state

and society, citizens became adept at constrained escalation, normative morality plays, and hiding the degree to which protests had leaders or organizational capacity to avoid the most severe repression (O'Brien & Deng, 2015; Fu, 2017).

The Chinese state has not eliminated repression from its tool kit during this period of responsive authoritarianism. Instead, repressive tactics have become more targeted, hidden, pre-emptive, and psychologically sophisticated (Ong, 2018; Chen, 2017; Cai, 2008). Deng and O'Brien examine the practices of "psychological coercion" and "relational repression" as strategies deployed by state agents who find it harder to dominate through sheer physical coercion or have lost the ability to withhold resources, especially to citizens with their market power. An aggrieved citizen might discontinue resistance at the insistence of colleagues or relatives (O'Brien & Deng, 2015; Deng & O'Brien, 2013). O'Brien & Deng (2017) note that "protest control is taking on a person-by-person quality" made possible by the vast increase in state resources dedicated to internal security and stability (see also Wang & Minzner, 2015; Ong, 2018).

China's progression toward more sophisticated repressive strategies is not unique. In many parts of the world, the use of non-physical, hidden, or delegated repression is widespread (Earl, 2003). Davenport (2007) shows how high capacity states can disable social movements without overt violence. Frantz & Kendall-Taylor (2014) find that the ability to co-opt potential opposition leaders can alter an authoritarian government's need for widespread repression while enhancing their ability to target opponents of the regime. Sophistication and "person-to-person" contact are not cheap. Observers noted in 2011 that China's internal security budget exceeded its spending on the military. Wang & Minzner (2015) detail the "rise of the security state" and its multiple tasks of bargaining, compensation, surveillance, and repression.

Escalation incentives are baked into the model of responsive authoritarianism (Tanner, 2004; Chen, 2012; Cai, 2010; O'Brien & Deng, 2015). This is evident from a pithy saying used by protesters: "Do nothing and get no results, do something small and get small results; do something big, and get a big result" (Tanner, 2004). With the rise of social media, escalation has become far easier with lower costs and more immediate impact. Online discussions can aggregate narrow and limited grievances (e.g., environ-

mental degradation, food safety, or lapses in public safety) into encompassing demands for greater transparency, a freer media, or bottom-up accountability for political elites. Netizens with large followings and influence on social media can rapidly bring these encompassing demands into the public discourse.

Responsive authoritarianism has also clearly benefited the government. Since 1989, it has gained legitimacy from its more accommodating stance toward protest and grievances (Dickson, 2016; Tang, 2016). Social media also provides the central government easier access to "real" public opinion, solving the "dictator's dilemma" of not having a true sense of how people feel about a repressive regime (Wintrobe, 2000). If people feel like they can be more honest in social media, the regime also benefits informationally. How can the government have its cake and eat it too? How can it sustain high levels of protest and response/bargaining within the context of extensive social media use? Can responsive authoritarianism survive the onslaught of virality?

2   Empirical Analysis

We draw upon two sources of data: 1) government documents and manuals about "opinion guidance," and 2) leaked internal documents from a social media company, Sina Weibo. The former serves as transcript evidence: a record from the government of its logic of control. The latter empirically tests our theory of information control using a model of the decisions of a popular Twitter-like social networking site, Sina Weibo, to escalate users and content on their platform to relevant authorities.

These leaked internal censorship documents help us test our theory of what kind of online content and users are relayed back to the government by media companies. It is evident from government documents and manuals analyzed in this paper that government outreach to social media "thought leaders" involves a combination of carrots and sticks.

Evidence from our analysis suggests that the state is most worried about controlling the virality of content, the influence of individuals, and the ability of social media discussion to mobilize and incite. In order to reduce the virality of a discussion or shape the course of a discussion, the state eliminates certain types of posts, reduces the influence of opinion

leaders, and gives a freer rein to the "nobodies" (for example, by allowing small users more leeway than big users).

2.1 Evidence from Government Documents

In this section, we present transcript evidence of the state's logic of information control: to focus repressive activities on influential users rather than categories of content. This section draws on data from dozens of Party- and government-produced books, documents, or manuals about "opinion guidance," "thought work," and information control, whose intended audience is government leadership cadres. The authors of these sources are high-level propaganda and public security officials, as well as government-affiliated academics.

Government logic of opinion guidance described in these manuals is consistent with our theory of "who not what." These documents show that the state focuses its opinion guidance efforts on users with large followings or viral posts: "opinion leaders" (yijian lingxiu, 意见领袖). Most opinion guidance manuals included at least a small section devoted to managing these "opinion leaders." Below, we will present this logic using excerpts from government manuals and documents.

Manuals suggest that the Party is concerned that it has ceded too much discourse power to voices outside the system. One manual suggests, "the mechanism for public opinion formation online has changed... [as the] means of digital dissemination has broken the [state's] traditional monopoly of discourse." This manual laments that social media has "brought about a change of position from strong to weak among elites, and from weak to strong among the grassroots." It suggests, "because cyberspace has no systemic barriers or binding ideological constraints, all kinds of thoughts, ideas, and values have a platform. Different classes, areas, and types of media can exchange, integrate, or confront these ideas, making the public opinion environment increasingly complex" (B. Zhou, 2011, 118-122). In excerpts presented later in this section, the state appears fearful of opinion leaders' ability to organize individuals around counter-hegemonic ideas. They suggest that in order to confront the expanded space for counter-hegemonic discourse brought by the internet, cadres must engage with opinion leaders who occupy the most social real-estate in the online opinion field. Manuals direct cadres to pay particular attention to

"'Big V' users who have many fans and followers... [because they] decide the development of online opinion events or the formation of topics of discussion" (Zeng, 2015, 180-194).

Manuals not only consistently stress the importance of opinion leaders in the state's opinion guidance efforts, but they also agree about the metrics cadres should use to track and identify them. One manual suggests the importance of two key metrics: repost counts and fan/friend counts. This manual suggests, "Weibo's special model for information dissemination makes it much easier for social media surveillance workers to identify opinion leaders and the original source of opinions" (B. Zhou, 2011, 201-228). Another manual elaborates on the logic behind targeting users based on these variables, citing internal government research on Weibo:

> "[In] over 30 viral opinion incidents on Weibo between 2011 and 2012, there were only 7584 viral posts where the number of retweets surpassed 500; a mere 305 Weibo users authored 5047 of these posts. Moreover, these 5047 posts [were the root source of] 66.5% of all total posts and 80% of all retweets and comments. Though the public opinion space seems complex and untenable, in reality, it is largely controlled and led by a small number of big V users..." (Zeng, 2015, 180-194).

In this same manual, the authors suggest that the state should balance their dual goals of responsiveness and control by regularly approaching opinion leaders, guiding them toward consensus and learning from their perspectives:

> "We must communicate with them, seek consensus, and let them make suggestions from a rational and constructive perspective, instead of publishing aggressive and inflammatory extreme speech as they please [...] it is necessary to maintain close contact with them, through regular or irregular meetings or online exchanges, to take the initiative to invite them to visit major public projects..." (Zeng, 2015, 180-194).

The manuals suggest that the state is keenly aware of the agenda-setting power and social import of celebrities, intellectuals, and artists online (Dang, 2013, 115-123). These individuals can "set a new agenda to lead the public to focus their attention elsewhere, resulting in a change in the overall direction of public opinion" (B. Zhou, 2011, 201-228). One manual discusses the case of a 2010 Shanghai high-rise fire that killed 58 people. It suggests, "changes in the agenda of discussion of this event were closely related to the actions of opinion leaders." In the aftermath of the disaster, celebrity blogger Han Han

(韩寒) and other famous journalists brought to light systemic issues behind the high-rise fire on social media platforms. Public outcry reached a fever pitch after a popular blogger organized a "public mourning event that attracted 100,000 participants in Shanghai." This protest was organized around systemic problems articulated by Han Han (B. Zhou, 2011, 201-228). The state, in turn, increased pressure on Han Han, who has since ceased his once prolific and acerbic blog writing. At the end of his blogging career, he famously remarked that "influence belongs only to those with power... they own the theater, and they can always bring down the curtain, turn off the lights, close the door, and turn the dogs loose inside" (Osnos, 2014).[1]

In order to prevent opinion leaders from fomenting challenges to the state, manuals stress the importance of "institutionalizing specially selected opinion leaders" as well as "discovering, cultivating, training opinion leaders, and encouraging leadership cadres to act as opinion leaders" (B. Zhou, 2011, 201-228). One source stresses that opinion guidance should prioritize co-optation rather than repression, even in cases when users are posting inflammatory information, a concept with revolutionary CCP origins: the "united front" (tongzhan, 统战). As Schurmann notes in the second edition of "Ideology and Organization," united front tactics take precedence when the organizational dominance of Communist Parties is not sufficient, as is the case in the online public sphere (Schurmann, 1968, 528). One manual calls on cadres to "take advantage of united front work with new media individuals" by "unifying with those outside of the party and outside of the system," "supporting those with good political values, who are familiar with online rules, and who are influential and popular 'big V' users" (T. Zhou, 2012). One manual furthers:

> "Develop an online 'united front' [...] do not fight. Look instead for areas of greatest agreement. Maintain the relative independence of opinion leaders, seek common ground on important issues while reserving differences on minor ones, gather and co-opt those with different points of view, deal with individuals on a case-by-case basis, support [those with the] 'correct' [opinion] and repress [those with the] 'wrong' [opinion]" (Zou & Su, 2015, 88-89).

Another manual similarly stresses that the state should not only focus on co-opting and persuading of opinion leaders but that it should tolerate and learn from a plurality of views:

---

[1]See http://www.webcitation.org/74wItIZaz.

> "When we identify ["grassroots"] opinion leaders on Weibo, we must strive to guide them in the right direction through [opinion leaders with the correct opinion]. In particular, we must pay attention to the role of network opinion leaders who are close to the identity of netizens, encourage their positive suggestions, and tolerate their radical speech" (B. Zhou, 2011, 201-228).

There are seldom content-based mechanisms suggested in these documents for monitoring online speech. Instead, social metrics such as repost counts and fan/friend counts are key inputs to "public opinion early warning systems." Public opinion emergency management reports published by the Chinese Academy of Social Sciences suggest this approach to identify "hot topics" using aggregate search volume and other measures of public attention. These reports suggest that the virality of a post and the influence of its author are more important criteria than content in the state's systems for pre-empting the spread of "harmful information." One manual confirms this, suggesting, "using suitable technological methods, real-time monitoring of information with high repost counts can be implemented to contain crises at the first sign of danger" (B. Zhou, 2011, 201-228).

2.2 Evidence from Sina Weibo Censorship Logs

To supplement transcript evidence from government sources, we draw on leaked censorship logs from social media company Sina Weibo. These data allow us to test our theoretical claims with records of censorship implementation at a social media company.

A whistle-blower at Sina Weibo shared these data with the Committee to Protect Journalists (CPJ), and they have been made available by CPJ at our request.[2] These data are logs of censorship implementation that record government directives from various bureaucracies to remove content, report statistics during crises, and inform on individuals to the government through Beijing-based government affairs liaisons. Each log can include information about users who have been reported to higher levels, quotations or paraphrased content the government wants to have removed, reports on changes in the perceived level of government monitoring, or the general political situation at a given time. In the logs, managers include decisions and instructions on censorship to be used by Sina's content moderators. These logs are used to share information between employees working on different shifts.

---

[2]The source has consented to the use of these data for our research and for replication purposes.

Logs sometimes give instructions to employees to "report up" or "report data" to government affairs liaisons in Beijing who cooperate with supervising Beijing propaganda and public security bureaucracies. Although we cannot wholly disambiguate between "reporting up" to facilitate repression and "reporting up" to facilitate co-optation, we know online speech and activity is often invoked by state security agents when they engage ("drink tea") with citizens. Reports by NGOs show online activity can be the starting off point for government repression, through "drinking tea" visits, arbitrary detention, and arrest (Tager, Glenn Bass, & Lopez, 2017). However, our analysis only covers the first step of reporting back to the government.

In total, there are 8,427 individual logs in our data. Logs are notes related to government directives, management censorship decisions, work guidelines, employee duties, and other administrative instructions. Logs disseminate management decisions about how/if to implement government directives to the large team of content moderators employed by Sina. A log will usually include descriptions or excerpts of content along with instructions to employees on how to proceed with content moderation. While many datasets related to censorship capture only keyword-based censorship or manual content review, these data include logs about both. This dataset includes the complete set of logs from 2011-2014. For this paper, we analyze only logs from 2012. We do so to limit the analysis to a single year under the Hu-Wen Administration. In other research, we explore the full range of data and the changes in censorship behavior in the early Xi Jinping administration (Miller, 2018).

2.2.1 Limitations and Scope Conditions of Logs

Because these logs are from a single social media company, there are limits to the external validity of our inferences. We do, however, have reason to believe our inferences can generalize to other social media companies, particularly Tencent (tengxun, 腾讯). Several logs indicate that competitor Tencent receives the same directives from the State Council Information Office, Shenzhen-based regulators, and the Cyberspace Administration of China (CAC). Tencent, like Sina, delays implementation, or disobeys directives when doing so gives them a competitive edge (Miller, 2016). Additionally, the market capitalization of both Sina and Tencent were similar at the time these logs were written,

which means both had similar leverage when it came to negotiating government directives.

The logs indicate that tech companies in China cooperate and influence each other's censorship decisions. For example, a log from May 7, 2013, discusses how Sina Weibo cooperates with Baidu (baidu, 百度) when developing keywords for content filtering.

Even a conservative approach to inferences drawn from these logs does not diminish their significance. Sina Weibo is a large and popular social media company in China. During the time of the logs, it was the most popular microblog and was ranked in the top 3 of all domestic social media companies by monthly active user statistics. During the time of the analysis, Sina Weibo boasted over half a billion registered users.

Though we can never fully verify the integrity of the source of the leak, we have been assured by the source and CPJ that the logs comprise a comprehensive set of internal orders of Weibo censorship departments between 2011 and 2014. Though we are confident in the integrity of the data, our additional analysis of state discourse in "opinion guidance" manuals and documents guards against the possibility of results that are confounded by idiosyncrasies of the log data.
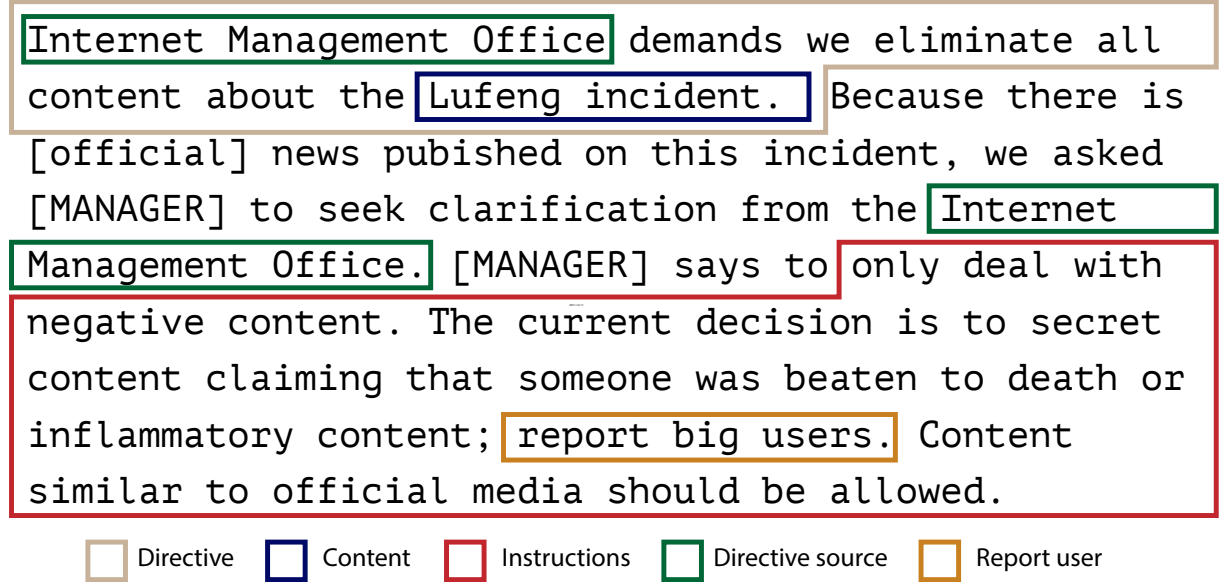
2.2.2 Content-coding and Variable Definitions

We code each log according to several content and instruction variables defined in Table 1[3]. Content variables reference the category of content targeted for censorship. Instruction variables refer to instructions on how to handle content rather than the topic category itself. Instructions refer to conditional statements in the logs. For example: "delete content *that is attacking the Party and government*" or "report users *with more than 10,000 followers*." See Figure 1 for an example of a censorship log and the coding of the log's raw text. The three content variables used in this analysis—collective action, government criticism, and local government corruption—are drawn from findings in the existing literature. King, Pan, & Roberts (2013) argue that the Chinese state tolerates government criticism while targeting threats of collective action. Dimitrov (2017) and Lorentzen (2014) argue that the government allows for social grievances to be aired for higher levels of government to collect accurate information about their lower-level agents.

---

[3]See code-books, diagrams, and inter-coder reliability measures in the appendix

We calculate instruction variables using inductively-developed keyword dictionaries. See detailed descriptions of variables used in this analysis in Table 1.

Figure 1: Example of a Censorship Log and Variable Types

Internet Management Office demands we eliminate all content about the Lufeng incident. Because there is [official] news pubished on this incident, we asked [MANAGER] to seek clarification from the Internet Management Office. [MANAGER] says to only deal with negative content. The current decision is to secret content claiming that someone was beaten to death or inflammatory content; report big users. Content similar to official media should be allowed.

☐ Directive ☐ Content ☐ Instructions ☐ Directive source ☐ Report user

2.3 Empirical Strategy

We measure the correlation between content and instruction variables and the reporting up of users to the authorities with a logistic regression model of log data from 2012. According to our theory of "who not what," we do not expect content variables to correlate with "reporting up" and instead expect attributes of the user and perceptions of virality to correlate with "reporting up." Though content *does* matter in the state's information control efforts, we argue the state intervenes in information control efforts when a user is influential, or content has gone viral. These attributes are strongly associated with whether Sina will report a user to the authorities. The model's dependent variable captures whether logs instruct employees to report individual users to the authorities.

In addition to our regression model, we illustrate these relationships using high profile cases from the logs. These cases are broadly representative of our interpretation of the dynamic between commercial censors and the state, but they vary in their level of political sensitivity and potential for collective action. The first case is a viral story about the citizenship of a high profile journalist. This case is less politically sensitive and has little collective action threat. The second case examines a civil society activist attempting to

Table 1: Variable descriptions

Variable	Description	Type
User Attributes	Any differentiation or categorization of users based on attributes such as fans and following counts, big/small user, media user, foreign user, verified user, etc.	Instruction
Retweets	Any mention of the number of reposts/retweets of a log.	Instruction
Govt. Attack	Any differentiation in instructions related to an attack on government officials, policies, or institutions.	Instruction
Political Humor	Any differentiation in instructions related to mocking, satirizing, or ridiculing government officials, policies, or institutions.	Instruction
Extreme	Any differentiation in instructions for the tone of content. Whether it is "ordinary" or "normal" vs. "abnormal" or "extreme." Note: this does not refer to content but rather to how it is framed or expressed.	Instruction
Rumors	Any differentiation in instructions that relates to rumors or fake news.	Instruction
Retweets	Any differentiation in instructions based on retweet counts or that relates to other attributes of post virality such as the number of times it is sent in a private message.	Instruction
Inciting	Any differentiation in instructions that relates to content that "incites" or "provokes." Though content about collective action is often "inciting," this category refers to sensational content, or that could bring about strong feelings of outrage.	Instruction
Govt.	The content mentions or implies a Chinese government institution, organization, or bureaucracy, a Chinese government official of any rank or position, their family members or their partners/mistresses, a Chinese government policy, or a Chinese state-owned enterprise (SOE).	Content
Local Govt. Corruption	The content mentions or implies any of the following: 1) misuse of local government office or local government funds, 2) sexual misconduct of local government officials 3) a local government official, and/or his/her family financially benefiting from a government post.	Content
Collective Action	The content either 1) mentions or implies an event where a group of people took action together to achieve a common objective, or 2) mentions or implies an individual or group of individuals who are advocating on behalf of a social, religious, or ethnic group.	Content

mobilize supporters to join his movement. This case is politically sensitive and includes explicit calls for anti-regime collective action. Finally, the third case –the Bo Xilai (薄熙来) incident– is a highly sensitive political case of intra-elite factional struggles, with low potential for collective action.

We examine how the content and instruction text in logs are associated with the decision to "report users," or send information about an individual user to a Beijing-based government affairs liaison who is responsible for coordination with provincial-level bureaucracies, the State Council Information Office, or the Cyberspace Administration of China. We present the results of our model in Figure 2 and a regression table in the appendix.

Figure 2: Logistic Regression Coefficient Plot
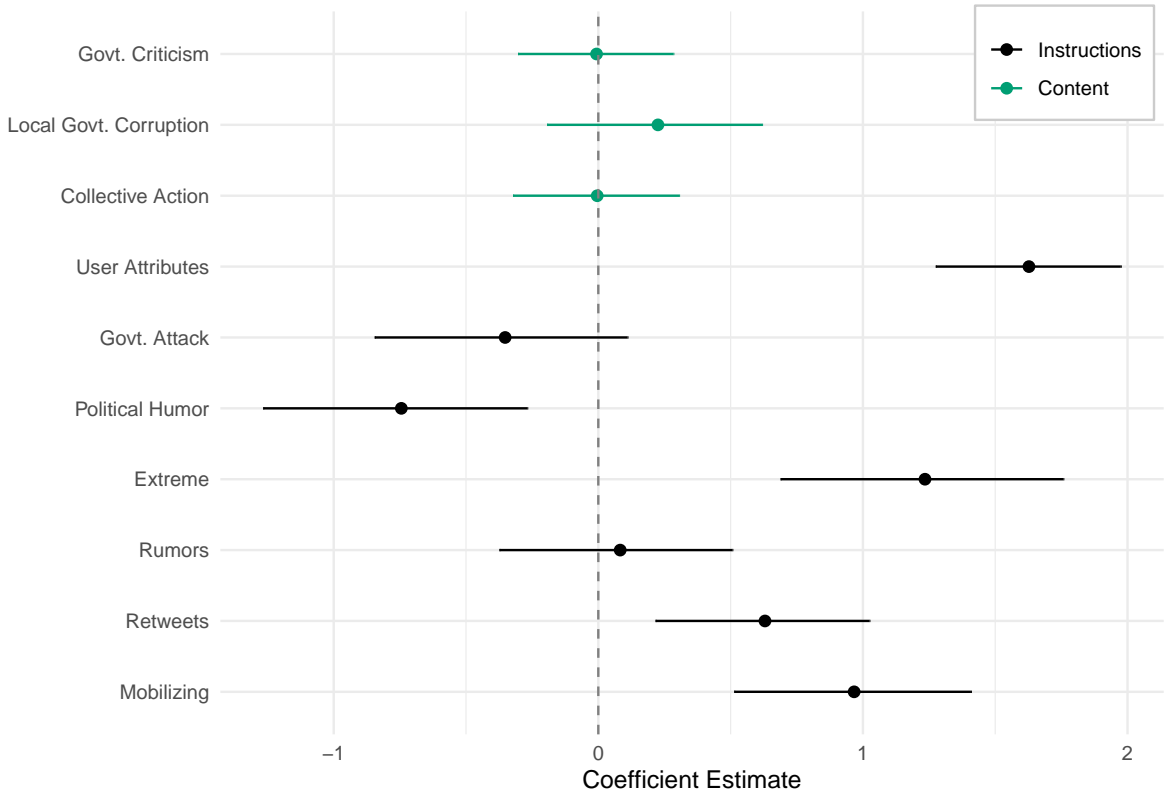
The biggest predictors of whether or not content or users are "reported up" are the presence of words that describe the influence of the user mentioned in the log (User), and, to a lesser degree, how "extreme" the content is (Extreme). This would include content that is making exaggerated moral and emotional appeals or advocating extremist positions. Mentions of retweets and the degree to which a post "incites" (shandongxing, 煽动性) are also significantly and positively associated with whether a user is "reported up." As we discuss in the cases below, this category would include posts that are inciting, even if only abstractly. It could include explicit exhortations for real-world social mobilization, but only if these exhortations incite or stir up strong emotions. By contrast, measured and journalistic statements of fact about collective action incidents would not fall into this category.

No content-based independent variables were significant. The censors instead appear to focus on the framing of content, e.g., content that "incites" rather than simply content about collective action events. For example, during an August 2012 taxi strike in the city of Hangzhou, the censors note, "Hangzhou taxis are on a collective stoppage today,

protesting the rise in gas prices and the state of the roads. This news can be circulated, but any posts that incite or appeal to other cities should be "made secret."[4] Report the original poster." Similarly, in logs about the 2013 arrest of Zhang Anni, the 12-year old daughter of a Tiananmen activist who wrote a letter to Barack Obama, the censors allow for discussion of the case, but not "incitement":

"Netizens expressing support for the Zhang Anni incident in Hebei, deal with any inciting content on Weibo, there is no need to delete general discussion at this point."[5]

— Log entry from April 10, 2013

The censors may realize that these are hot button topics, and deletion is not in their commercial interest. It is also possible the state itself is tolerant of more open discussion because while some will sympathize with the striking taxi drivers or young activists, others will criticize them for inconveniencing customers or being insufficiently patriotic. Han (2018) argues that the state benefits from a more open online atmosphere because this allows "discourse competition" between supporters and opponents of the government. These posts reveal that thin line tread by the commercial censors. Sina tolerates discussion of collective action and political dissidents up to a point but reports it to the authorities if it "incites."

The commercial censors report up information about influential "big V" users, and viral, inciting news. Online information control is deployed alongside real-world physical repression for the die-hard activists, as we see below in the case of Xu Zhiyong (许志永), the civil society activist.

2.4   Selected Cases

In the section below, we discuss three emblematic cases of the phenomenon of "reporting up." These cases are selected because they represent three areas of government concern: government legitimacy (Yang Lan's citizenship), political dissidence (Xu Zhiyong's arrest), and elite defection (Bo Xilai's trial). Only the second is a clear case of a call for collective

---

[4]A type of "shadow-banning," a censorship tactic where posts are hidden from a subset of users rather than deleted. When a post is "made secret," the original poster can see it, but no one else can.

[5]Original Chinese: "网友声援合肥张安妮事件，微博中处理煽动行动内容，一般讨论的暂时不必处理。"

mobilization against the state. As such, these cases demonstrate the significance of user characteristics across diverse areas of political sensitivity.

2.4.1 Yang Lan's Citizenship

On March 7, 2012, the censors at Weibo were getting off the night shift. As they wrote up guidance and instructions for the next shift, they wrote, "Saying that Yang Lan is a US citizen has already been declared a rumor, directly delete posts of small users, report up big users."[6] A few days later, they repeated the command, "[Suggesting that] 'Yang Lan is a US citizen' should be considered fake news, if it is seen on Weibo, delete it, report big users."[7]

The censors were attempting to shut down a raging debate on Weibo, "Is American citizenship still valuable?" On March 6, there were more than 1.5 million posts on the topic. Some of the discussion focused on the post-Global Financial Crisis state of the US economy, but other Netizens took aim at their compatriots. "If you say China isn't good, you are a Western slave. If you say America is good, you are an American bitch. If you say you don't want to be Chinese, you are a wretch and a traitor [...] If you get a green card and stand on the corner of American streets shouting, 'I love you, China,' you are a 'patriot.'[8]

Yang Lan (杨澜), a celebrity journalist and TV host, "China's Oprah Winfrey," was alleged to be this kind of person. She appeared patriotic and had a strong political background (her father was a translator for Zhou Enlai), but after making a fortune in China's media market, she had taken citizenship elsewhere. Yang tried to rebut the accusations. Later, she and her wealthy husband, Bruno Wu, threatened to sue.

In the Yang Lan case, the censors treated users differently based on how important and influential they are. Sina deleted normal users' posts when they gossiped about or criticized Yang Lan; they deleted the accounts and reported opinion leaders who did the same. While collective action may have been a concern in this case, it is more likely that

---

[6]Original Chinese: 说杨澜是美国籍的已经辟谣了，小用户直接删除重要用户报一下

[7]Original Chinese: 传杨澜为美国籍为虚假消息微博见到删除重点用户报一下.

[8]Archived Atlantic article available here

the government was concerned about its legitimacy. If someone like Yang Lan, who is well-connected, wealthy, successful, and "within-the-system" is seeking foreign citizenship, it reflects poorly on the CCP. Are elites losing faith in the country's future? Do even the well-connected and wealthy have an exit option should the future take a turn for the worse? The crackdown on the Yang Lan case shuts down these questions.

2.4.2 The Arrest and Trial of Xu Zhiyong

In cases involving well-known political activists, however, the concern with influence is more directly connected to social mobilization and collective action. For example, Xu Zhiyong, a well-known political activist and liberal reformer, first became famous in 2003 as part of a group of intellectuals who challenged the state's "custody and repatriation system" that detained rural migrants who did not have legal documentation to live in cities. He later opened an NGO, Open Constitution Initiative (Gongmeng, 公盟), which championed the rights of migrants to education in cities and investigated other scandals. He was detained in 2009 and Gongmeng was accused of tax evasion. At the time, Xu was working on the case of the tainted milk scandal in 2008, which had led to the deaths of several infants who consumed adulterated baby formula. In May and August of 2012, Xu Zhiyong used Weibo to ask for citizen support for Gongmeng, even soliciting donations online and calling for unity among Chinese citizens. Both of the posts below were instructed to be "made secret," "big V" users who posted such material were reported up.

"On May 23rd, Xu Zhiyong used domestic sites (Sina Weibo and Tencent Weibo) and foreign sites (Twitter) to incite Netizens to join *Gongmeng*, to purchase a "citizen" badge. Even calling for people to sign a "citizen promise." [...] If this content is seen on Weibo, make it secret. Report big users to the authorities."[9]

— Log entry from May 26, 2012

---

[9]Original Chinese: "许志永"于5月23日在境内《新浪微博》、《腾讯微博》及境外《推特》发布信息煽动网民加入"公盟"组织，购买"公民"徽章。并呼吁网民签署《公民承诺》。原文称："需要公民徽章的请给我发信[EMAIL]，我们需要再订制。其实不需要传统意义上的所谓加入，公民（公盟）是自由公民的联合，我们都是时代新公民，为公民社会共同努力！请发信至[EMAIL],常联系。这个内容微博中见到私密，重要用户发的上报负责人"

"If 10,000 citizens each give *Gongmeng* 100 yuan, we could do so many things! We earnestly request support for *Gongmeng*; please believe we will use the money for the places that are most in need of justice. Legal aid assistance account: Xu Zhiyong, China Industrial Bank Beijing Branch, [BANK INFO] - Secret posts calling for donations to *Gongmeng*."[10]

— Log entry from October 8, 2012

In 2013, activists, including Xu Zhiyong, called for greater disclosure of government salaries and wealth. Xu Zhiyong championed this initiative as part of his new venture, The New Citizens Movement.[11] Demonstrations were held in several cities and led to the arrests of many people. After being under house arrest for three months, Xu Zhiyong was finally detained in July 2013, charged with "attempting to disturb public order" in August, and sentenced to four years in prison after a trial in January 2014. (He was released from prison in July 2017.)

During this time of renewed activism around transparency issues, the commercial censors were given many orders to censor information about Xu Zhiyong and, importantly, to report up "big users" and "special circumstances." Certain users were named in particular, including Xu's defense lawyer, Liu Weiguo, and other rights defense lawyers. On the same day of the post below, Liu was himself detained:

"Regarding the issue of several lawyers going to Beijing's Third Detention House to visit the detained Xu Zhiyong, when verifying, make [this information] secret. Increase the number of keywords that are automatically "made secret." Immediately report up any special circumstances, including the lawyers, Chen Jiangang (陈建刚), and Liu Weiguo (刘卫国). These two say a lot. While verifying, pay careful attention; these words have already been verified."[12]

— Log entry from July 18, 2013

With well-known political activists who openly challenge the government several fronts, the repressive apparatus of the state is not hidden or subtle. The commercial censors are told the issue is of great importance, and online discussion is used to monitor particular

---

[10]Original Chinese: "如果有一万个公民每年每人捐给公盟100元，我们能做多少事啊！恳请支持公盟，请相信我们会把钱用到最渴求正义的地方。法律援助救助账号：许志永，中国工商银行北京分行学院路支行，[BANK INFO]，支付宝兼恳请支持，望扩散号召给公盟捐钱的私密"

[11]A more detailed analysis of this movement is in Pils (2014).

[12]Original Chinese: "关于有几位律师和网民去北京第三看守所看望被关押的许志永一事，审核中见到都私密处理，加了部分自动私密关键词，有特殊的情况及时上涉及的律师：陈建刚刘卫国这两个人说的比较多，审核的时候多注意，已经是审核词了"

people, both virtually and eventually, through physical detention and criminal charges. Many of the lawyers mentioned in these posts were later detained in the July 2015 massive crackdown on human rights lawyers across China. Many other "rights defense" lawyers have gone underground since the arrests of 2014 and 2015 (Pils, 2014).

The Xu Zhiyong incident reveals the government's desire to shape and guide public opinion, not just repress the actions of these few or stop collective action. During the January 2014 trial of Xu, for example, censors were instructed to calibrate censorship carefully:

"Tonight at the First Intermediate Court in Beijing, Wang Gongquan (王功权) admits guilt that he and Xu Zhiyong incited people to disturb public order, he has deeply reflected and awaits trial. Related content can be discussed as long as it reflects the media reports. Make secret anything that does not reflect the official line, stop small users and small media from discussing. Discussions in the big media should be verified first, posts with lots of retweets should be carefully reviewed."[13]

— Log entry from January 22, 2014

Both the Yang Lan and Xu Zhiyong cases demonstrate the targeting of influential users as the regime seeks to control what is said and heard about China's elite. The cases are different in the sense that Yang Lan's citizenship is an abstract threat to the CCP's standing, while Xu Zhiyong's exhortation to fellow citizens is a concrete call for social mobilization. It is no coincidence that these cases involve a famous lawyer and even more famous journalist, part of China's new professionalized and cosmopolitan elite. It is critically important to the Party not only what they do or say online, but also what other people say about them.

2.4.3 Bo Xilai's Scandal

The importance of external elite voices was apparent during the 2012-2013 power struggle, and potential coup, which pitted Chongqing Party Secretary Bo Xilai and the incoming leader, Xi Jinping. As the scandal unfolded, influential users who expressed support for Bo were reported to the authorities. Additionally, the government demanded that Sina

---

[13]Original Chinese: "王功权承认违法犯罪被取保候审】北京市一中院今晚称，王功权承认与许志永策划煽动聚众扰乱公共场所秩序的违法犯罪活动，王功权表示深刻反省；对王功权予以取保候审。相关内容目前可以按照该新闻口径讨论，凡是与新闻不一致的私密处理，小用户小媒体禁止评论，大媒体评论先审转高审。"

Weibo monitor and report public opinion data at fixed intervals. In the log below, we see Sina implementing requests to report data about public opinion related to Bo Xilai. This log serves both an informational role and a repressive role:

"Handling Bo Xilai-related posts: Total posts censored, among them, support for Bo (#, %), ridicule [of Bo] (#, %), reposting foreign media (#, %), other (#, %). [Also report] whether or not there are any influential users directly supporting Bo. Also, today's Bo Xilai reporting should be calculated according to the following intervals. The reporting up schedule has been revised to the following three time slots: 10:00, 18:00, and 23:00."[14]

— Log entry from April 12, 2012

"[Keep track of the] number of deleted and filtered posts about Bo Xilai and Wang Lijun, the number of deleted and filtered comments, how many of these are articles oppose the way the central government has handled Bo Xilai, are unsatisfied with the results of the investigation of the case, how many are republished from foreign websites, and other related information. The reporting timeline has been changed as follows to these three times: 11:00, 16:00, and 23:00."[15]

— Log entry from April 13, 2012

The Bo Xilai scandal was what the Party calls a "sudden, unexpected emergency" (tufa shijian, 突发事件). As such, immediate public opinion responses were followed with data collection and planning for more comprehensive public opinion interventions as the situation developed. More targeted interventions required information gathering, which is why Sina was asked to report data to the authorities. In later stages of the crisis, more specific topics of conversation became off-limits, not because there were pre-existing procedures to target specific categories of content, but because the goal of "opinion guidance" is to reduce the visibility and influence of content propagated by counter-hegemonic voices and increase the visibility of state narratives. An example is a more specific log from the Politburo Standing Committee's inspection tour of Chongqing following the incident:

---

[14]Original Chinese: "有关薄的处理数XX条，其中，支持薄XX条，占XX%，调侃XX条，占XX%，转发外媒XX条，占XX%，其他情况XX条，占XX%。有无影响力大的用户明显直接挺薄内容。今天薄熙来的数同时按这个时段统计这个报数时间段改为10点、18点、23点三时段"

[15]Original Chinese: "薄和王的删除、过滤数，共删除条，过滤条，其中反对中央对薄熙来处理决定的篇；不满案件调查结果的篇；转载境外网站的篇；其他类信息篇。11点、16点、23点三时段以后时段更改如下。"

"Today [MANAGER] published content moderation instructions for the inspection tour of Chongqing by the nine members of the Politburo Standing Committee (PSC):

1. Manage all content (including news) related to the inspection tour of Chongqing by the nine members of the PSC.
2. Manage all rumors about government, the most important being: foreign reports and speculation, reports from foreign media where the size of discussion in the comment section is particularly large, content originating from sensitive individuals, commentary on Sina Online official accounts, and headline news.
3. For users who publish rumors, the Public Security Department will investigate and deal with the person at the time, and at the same time, investigate the criminal responsibility of the person in charge of the website.
4. In mid-May, government rumors were thoroughly cleaned up, and whenever one appeared, the owner of the website was required to go to a meeting at the city bureau and rectify their errors."[16]

— Log entry from April 27, 2012

Later, as Bo Xilai was tried in the Jinan Intermediate People's Court, the logs portray a more planned and carefully executed opinion guidance strategy. Many China watchers expressed surprise when the Bo Xilai trial was covered live on Weibo, but this was done carefully and with Sina Weibo's cooperation.[17] On the opening day of the trial, Weibo was instructed to not only target categories of off-limits content but aid the Party in its efforts to set the agenda by heavily controlling counter-hegemonic discourse from "Big V" users and various non-official media accounts:

"... All comments discussing the matter from big media users and big V users are to be reviewed first before publication, one hour after the broadcast, lock comments. Directly prevent comments on small media accounts."[18]

— Log entry from August 22, 2013

---

[16]Original Chinese: "今日[MANAGER]布置的九常委视察重庆的清理要求：1：有关常委去重庆视察的内容（包括新闻，薄出事之前，所有微博账户）全部私密处理。薄被处理的新闻通稿不要动。下周一前必须完成，做好分类的数据统计。常委名字+重庆、模式、唱红、打黑。2：以后的时政谣言，从严处理，对用户必要时候上手段控制用户。3：媒体评论，北京负责审核发现，及时报备，管理不住的常委新闻，就关闭评论，天津负责清理。以上要求：监控所有人员务必执行。通知原文：1：有关常委去重庆视察的内容（包括新闻）全部处理。2：时政类谣言全部清理，主要类型为：外站报道及猜测、境内媒体的报道后面的评论及尺度比较大的一些新闻内容、敏感人士发布的内容、新浪网官方账户的评论，如头条新闻。3：发布谣言的用户，公安部门将依法查处当时人，同时追究网站负责人的刑事责任。4: 5月中旬彻底清理干净时政类谣言，凡出现一条，网站负责人必须去市局开会，进行整改。"

[17]See http://www.webcitation.org/74k5wmpno

[18]Original Chinese: "所有讨论此事的大媒体评论以及大V的评论均先审后上，一小时后锁定评论，小媒体发表此事直接关闭评论。"

In several related logs on the topic, content moderators were instructed to report up influential users and hide all content that was not exactly as reported by the Jinan Intermediate People's Court's Weibo account. By restricting discussion to official channels and focusing primarily on influential counter-hegemonic voices, the state was able to appear open, transparent, and legitimate in its trial of the fallen Party Secretary.

In the handling of all three, Yang Lan's citizenship, Xu Zhiyong's trial, and the intrigue over the Bo Xilai affair, commercial censors delete or make secret information that mobilizes or incites public opinion, they report influential users to the authorities, and they also report up data on public opinion. Their tasks are multifaceted, and to focus only on the repressive activity misses the more sophisticated goals of guiding public opinion and delivering more accurate information to leaders about what opinion leaders and the masses are thinking regarding controversial topics. However, information about specific users can always be retained for future repressive activities. In an April 2011 log about a taxi and truck driver strike, the link between information and repression is apparent:

"Strengthen handling of inciting and harmful information about the taxi and truck drivers' strike against the rise in gas prices, especially information that mobilizes for the June 1 strike. Accumulate related news and retain the IP addresses (of users) with information that incites or mobilizes."[19]

— Log entry from April 15, 2011

In an interview with the source of the leak, he recalls, "when some political rumor was spread widely, or during the Tiananmen Square anniversaries, police harassed people based on the information we provided." The source recalls a manager once saying, "hand in your big form. The police need it to arrest people" when instructing employees at Sina Weibo's censorship office in Tianjin (Wang, 2016b,a).

3   Concluding Discussion

This paper contributes to our general understanding of the governance strategies of the Chinese Party-State in the virtual public sphere of social media. Building off earlier theories that emphasize the importance of limiting social elites external to the CCP, we find the

---

[19]Original Chinese: "网监通知：进一步加强借"油价上涨"煽动出租车及货运司机罢运、罢工等的有害信息的清理工作，特别是煽动6月1日罢工的信息。相关信息自行积累，行动性、煽动性信息查IP留存。以便随时统计。昨日清理，目前未发现相关行动性信息。"

government is concerned with the power of virality and influence on the Internet, a space that cannot be easily dominated simply via Communist Party organization. Attempts to mobilize citizens collectively via online discussions, such as in the Xu Zhiyong episode, are threats to the government's hold on power. However, the government attempts to limit "counter-hegemonic discourse" even when the threat of real-world collective action is not imminent. Work on social contention suggests that the Party may target influential thought leaders because they have large networks and can threaten information cascades that could undermine CCP legitimacy, encourage real-life collective action, or shift public opinion on important topics. Big-V users function as brokers by connecting social media users that previously may have had no connections. Large numbers of online followers create authority and discourse power that challenge CCP dominance.

Maintaining responsiveness and more accurate monitoring of public opinion are essential facets of the government's strategy, however. In a sense, these findings underscore the information problems that confront dictators (Wintrobe, 2000). Social media discussion is desirable because it elucidates public opinion trends and social grievances. Nevertheless, it is also too much of a good thing.

Going back to the exposition of the importance of Party leadership in Schurmann (1968), we confirm the CCP is most vociferously interested in constraining the rise of powerful alternative voices and influence. Schurmann's analysis is focused on the period when the old elite of the ancien regime is vanquished. Our focus is on a period when social media elevated new voices and allowed opinion leaders to re-emerge independent of the Party. Schurmann's perspective is still helpful, however, given its focus on the role of elites in leadership. The rise of the virtual public sphere created new challenges for the Party to maintain its discourse power and its organizational hegemony. As the newest battleground for the hearts and minds of citizens, social media platforms are critical spaces for the CCP to inhabit and dominate. New elite voices and opinion leaders must be co-opted or repressed.

The government is intolerant of many topics once they become viral or are circulated by influential opinion leaders. The content of the post is less important than who is posting it and how many people are reposting it. The discussion over Yang Lan's citizenship, for

example, seems to fall into this category. It does not immediately threaten any sort of collective action or revolt, but rather it is part of a gradual process in which the legitimacy of the Party is undermined and sabotaged by elites and influential thought leaders.

Schurmann's work is instructive in part due to its focus on the 1950s, the period immediately after the Communist Revolution when the CCP was constructing the building blocks of its rule, and the 1960s, in the second edition, when social mobilization during the Cultural Revolution undermined the CCP's organizational dominance. In the realm of today's social media, the CCP is similarly limited in its capacity to take over all the space afforded by the Internet, nor does it wish to do so given the advantages of having more free discussion and better information about public opinion. Instead, the CCP focuses on repressing or co-opting voices that could threaten its position of leadership, to counter what Lei & Zhou (2015) call the "counterhegemonic public sphere." This is an arena in which citizens are exposed to and generate discussion often beyond the confines of the official government position. The critical threat is not the degree to which public opinion is representative, but rather how influential one is. They write:

> As the propaganda official put it, in the end, it is those who speak up instead of those who keep silent that influence other people and bring trouble for the Chinese government (Lei & Zhou, 2015, 455).

Despite the relevance of Schurmann's conceptualization of Party governance, it cannot fully capture the complex way in which the CCP manages virtual public spaces in a far more open, diverse, and vibrant context than experienced during the Maoist period. Moreover, more information and the "appearance of freedom" have real benefits to the government, both in terms of its legitimacy and its understanding of public opinion. Repnikova (2017) finds the authorities shape media policy by carefully scouring social media. She finds, "alongside control, Chinese authorities scrupulously listen to and study public opinion online, engage with and respond to public grievances, and creatively mobilize the public through interactive social media" (Repnikova, 2018). Stockmann (2013) shows that allowing for greater media liberalization with continued content control can bolster people's trust in the news. Han (2018) argues, "discourse competition" between regime supporters and opponents benefits the state by revealing that Netizens are not all liberal CCP opponents. There are true believers, emboldened by the anonymity and the

network-building available online. Finally, Roberts (2018) argues that censorship is more like a "tax" than a simple restriction on information. Most Netizens will never seek out sensitive political topics online or purchase a VPN to read foreign media. As these leaked blogs seem to indicate, the censors do indeed allow for a great deal of discussion online to happen while systematically targeting those few users who have the influence and the public following to cause real damage.

Appendix

Logistic Regression Results

Table 2: Logistic Regression Model for "Reporting Up"

	Dependent variable:
	Report Up
Extreme	1.234***
	(0.272)
Govt. Attack	−0.352
	(0.243)
User Attributes	1.627***
	(0.178)
Rumors	0.083
	(0.224)
Retweets	0.630***
	(0.206)
Mobilizing	0.967***
	(0.228)
Political Humor	−0.744***
	(0.253)
Govt. Criticism	−0.007
	(0.150)
Local Govt. Corruption	0.225
	(0.207)
Collective Action	−0.004
	(0.160)
Observations	2,212

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

Content Variables

Figure 3: Collective Action Coding Diagram
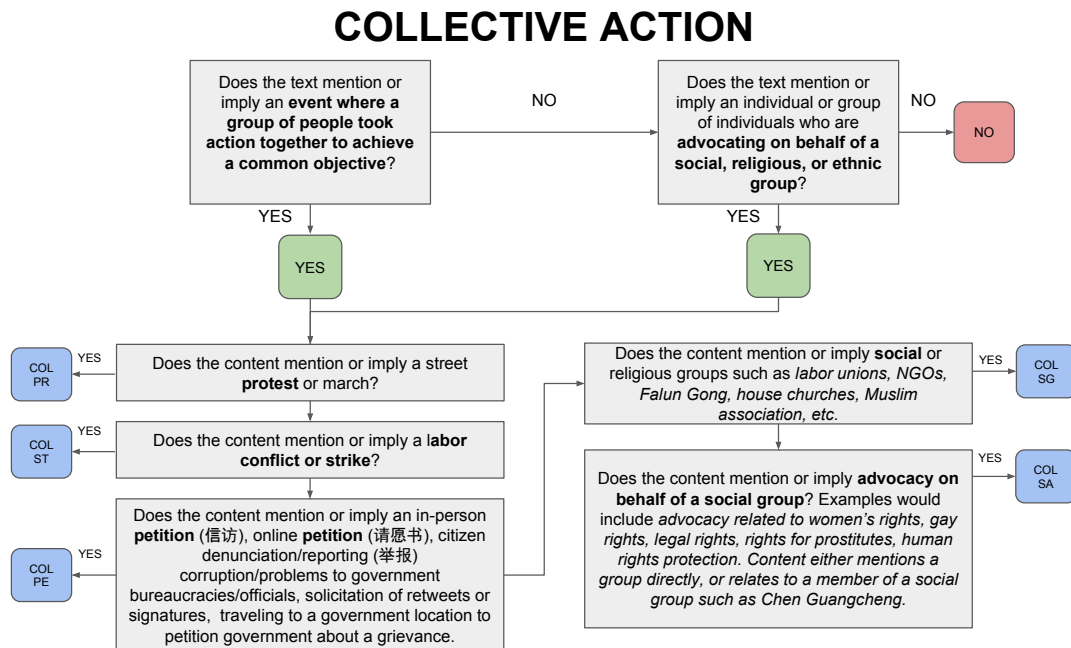
COLLECTIVE ACTION

Figure 4: Corruption Coding Diagram
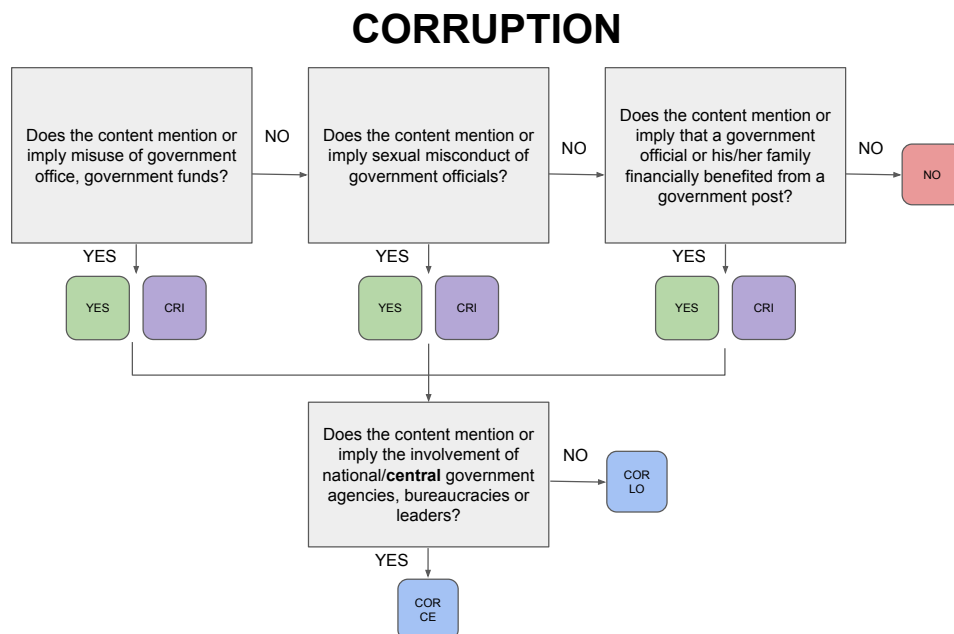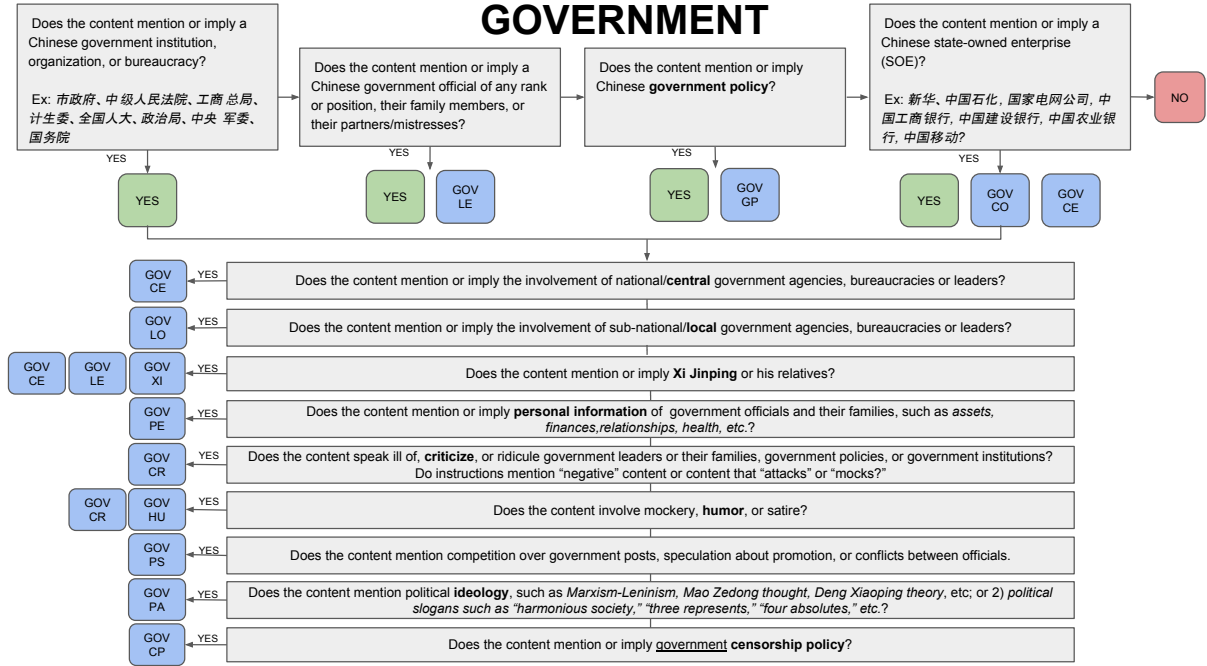
CORRUPTION

Figure 5: Government Coding Diagram

Instruction Variables

3.1 Report Users

The instruction text contains a (near) match of the following phrases used about a user (用户, UID):

- Report (报备, 报一下)
- Report up (上报)
- Send to Beijing (发给北京)
- Send it to the authorities (发给负责人, 提供给负责人)

3.2 Report Data

The instruction text contains a (near) match of the following phrases used about content (信息, 内容, 文章号, 言论, 评论):

- Report (报备, 报一下)
- Report data (报数)
- Report up (上报)
- Send to Beijing (发给北京)

- Send it to the authorities (发给负责人, 提供给负责人)

3.3 User Attributes

The instrucation text contains a (near) match of the following attributes of users (用户, 账号, 号):

- Small [user] (小[用户,号])
- Big [user] (大[用户,号])
- Big V user—verified user with many fans, V—verified user, VIP (大V, V, VIP)
- Opinion leaders (意见领袖)
- Famous people (名人)
- Media [users] (媒体)
- Grassroots (草根)
- Important [users] (重要)
- Ordinary [users] (普通)
- Delete for ordinary users and make secret for verified users (普删V私)
- [Users] with many fans (粉丝较多的)
- [Users] with around [#] fans (粉丝数[#]左右的)
- Newly-registered [accounts/users] (新注册)
- Old accounts (老账户)
- Leftist [users] (左派)
- Rightist [users] (右派)
- [Users] who post memes (段子类)
- "Reincarnated" [users]—users who continually create new accounts after deletion (转世)
- [Users] who continue to post (连续发布的用户)

3.4 Extreme

The instruction text contains a (near) match of the following phrases used about content (信息, 内容, 文章号, 言论, 评论):

- Extreme (过激的)
- Excessive/overboard (过分的)

- Vicious/malicious (恶毒的)

- Malicious (恶意的)

- Disgusting/gross (恶性的)

- Disgusting/abominable (恶劣的)

- Inappropriate (不行的)

- Attacking (攻击的)

- Hurling invective (谩骂的)

3.5 Attacking the Government

The instruction text contains a (near) match of the following phrases used about content the country, the party, the internet management system, the party, party-state, country, government, system, social system, leadership, the Central Propaganda Department, Xi Jinping (我国, 我党, 我互联网管理制度, 党, 党政, 国家, 政府, 体制, 社会制度, 领导人, 中宣部, 习近平)

- Attack ＿＿＿ (攻击＿＿＿)

- Make a negative connection with ＿＿＿ (负面联系＿＿＿)

- Concerning something negative about ＿＿＿ (涉及＿＿＿负面的)

3.6 Rumors/Fake news

The instruction text contains a (near) match of the following phrases used about content (信息, 内容, 文章号, 言论, 评论):

- Fake (虚假)

- Untrue information (内容不实)

- Spreading fake news (传播虚假消息)

- Spreading rumors (造谣)

- Unconfirmed true or false (不确定真假)

- According [to rumors] (传)

- Refuting rumors (辟谣的)

3.7 Retweets

The instruction text contains a (near) match of the following phrases used about content (信息, 内容, 文章号, 言论, 评论):

- Retweets cannot surpass [#] (转发不能过[#])
- Retweets surpass [#] (转发超过[#]的)
- Lots of retweets (转发多的, 转发高的, 转发较高的)
- Every hour surpassing [#] retweets (每小时总量超过[#])
- Pay attention to retweets (注意转发)
- Severely restrict retweets (严禁转载)
- Negative retweets (负面转发)

3.8 Inciting

The instruction text contains a (near) match of the following phrases used about content (信息, 内容, 文章号, 言论, 评论):

- Inciting (煽动性)
- Mobilizing (行动性)
- Collective (群体性)
- Call on/appealing to (号召的)
- Cause a fuss, make a scene (闹事)
- Causing a panic (造成恐慌的)

3.9 Political Humor

The instruction text contains a (near) match of the following phrases used about the country, the party, the internet management system, the party, party-state, country, government, system, social system, leadership, the Central Propaganda Department, Xi Jinping (我国, 我党, 我互联网管理制度, 党, 党政, 国家, 政府, 体制, 社会制度, 领导人, 中宣部, 习近平):

- Make fun of ＿＿＿ (拿＿＿＿开玩笑)
- Ridicule (调侃)
- Satirize (讽刺的)

- Meme/joke (段子)
- Satirical humor(恶搞)

Intercoder Reliability

These are the results comparing the predictive power of coder 1 to coder 2's decisions using the area under the receiver operator curve (ROC AUC). All categories are above the acceptable measure of .7 and many are above .85, which is highly reliable. We also prevent another similar measure, Krippendorf's Alpha, which also shows reliability of the relevant categories.

Table 3: Intercoder Reliability Measures

Category	% Agreement	AUC	Krip. Alpha
Col. Action	0.90	0.88	0.76
Corruption	0.95	0.93	0.86
Local Corr.	0.95	0.88	0.79
Govt.	0.90	0.84	0.69
Govt. Criticism	0.83	0.82	0.65

References

Cai, Y. (2008). Local governments and the suppression of popular resistance in china. *The China Quarterly*, *193*, 24–42.

Cai, Y. (2010). *Collective resistance in china: Why popular protests succeed or fail.* Stanford University Press. Retrieved from http://www.sup.org/books/title/?id=16868

Chen, X. (2012). *Social protest and contentious authoritarianism in china.* Cambridge University Press.

Chen, X. (2017). Origins of informal coercion in china. *Politics & Society*, *45*(1), 67-89. Retrieved from http://dx.doi.org/10.1177/0032329216681489 doi: 10.1177/0032329216681489

Dang, S. (2013). *Wangluo yulun hudie xiaoying yanjiu: Cong "weineirong" dao yulun fengbao.* People's University Press.

Davenport, C. (2007). State repression and political order. *Annu. Rev. Polit. Sci.*, *10*, 1–23.

DeNardo, J. (2014). *Power in numbers: The political strategy of protest and rebellion.* Princeton University Press.

Deng, Y., & O'Brien, K. J. (2013). Relational repression in china: using social ties to demobilize protesters. *The China Quarterly*, *215*, 533–552.

Dickson, B. (2016). *The dictator's dilemma: The chinese communist party's strategy for survival.* Oxford University Press.

Dimitrov, M. K. (2017). The political logic of media control in china. *Problems of Post-Communism*, *64*(3-4), 121–127.

Earl, J. (2003). Tanks, tear gas, and taxes: Toward a theory of movement repression. *Sociological Theory*, *21*(1), 44–68.

Frantz, E., & Kendall-Taylor, A. (2014). A dictator's toolkit: Understanding how co-optation affects repression in autocracies. *Journal of Peace Research*, *51*(3), 332–346.

Fu, D. (2017). Disguised collective action in china. *Comparative Political Studies*, *50*(4), 499–527.

Gao, H., & Zhang, Z. (2011). *Wangluo yuqing yu shehui wending.* Xinhua Press.

Han, R. (2018). *Contesting cyberspace in china: Online expression and authoritarian resilience.* Columbia University Press.

Hasmath, R., & Hsu, J. Y. (Eds.). (2017). *Ngo governance and management in china.* Springer.

Heurlin, C. (2017). *Responsive authoritarianism in china.* Cambridge University Press.

King, G., Pan, J., & Roberts, M. E. (2013). How censorship in china allows government criticism but silences collective expression. *American Political Science Review*, *107*(02), 326–343.

Kuran, T. (1991). Now out of never: The element of surprise in the east european revolution of 1989. *World politics*, *44*(1), 7–48.

Lee, C. K., & Zhang, Y. (2013). The power of instability: unraveling the microfoundations of bargained authoritarianism in china. *American Journal of Sociology*, *118*(6), 1475–1508.

Lei, Y.-W., & Zhou, D. X. (2015). Contesting legality in authoritarian contexts: Food safety, rule of law and china's networked public sphere. *Law & Society Review*, *49*(3), 557–593.

Lohmann, S. (1994). The dynamics of informational cascades: The monday demonstrations in leipzig, east germany, 1989–91. *World politics*, *47*(1), 42–101.

Lorentzen, P. (2014). China's strategic censorship. *American Journal of Political Science*, *58*(2), 402–414. http://www.peterlorentzen.com/research_files/Lorentzen_censor_2014.pdf.

Lorentzen, P., Fravel, M. T., & Paine, J. (2017). Qualitative investigation of theoretical models: the value of process tracing. *Journal of Theoretical Politics*, *29*(3), 467-491. Retrieved from https://doi.org/10.1177/0951629816664420 doi: 10.1177/0951629816664420

Meissner, M. (2017, May). *China's social credit system.* Retrieved from https://www.merics.org/fileadmin/user_upload/downloads/China-Monitor/merics_ChinaMonitor_39_englisch_Web.pdf

Miller, B. (2016). *The limits of commercialized censorship in china.* (Unpublished paper draft)

Miller, B. (2018). *Reassessing the targets of china's online censorship apparatus.* (Unpublished paper draft)

Nathan, A. J. (2003). Authoritarian resilience. *Journal of Democracy*, *14*(1), 6–17.

O'Brien, K. J., & Deng, Y. (2015). Repression backfires: tactical radicalization and protest spectacle in rural china. *Journal of Contemporary China*, *24*(93), 457–470.

O'Brien, K. J., & Deng, Y. (2017). Preventing protest one person at a time: Psychological coercion and relational repression in china. *China Review*, *17*(2), 179–201.

O'brien, K. J., & Li, L. (2006). *Rightful resistance in rural china.* Cambridge University Press Cambridge.

Ong, L. H. (2018). "thugs-for-hire": Subcontracting of state coercion and state capacity in china. *Perspectives on Politics*, *16*(3), 680–695. doi: 10.1017/S1537592718000981

Osnos, E. (2014). *Age of ambition: Chasing fortune, truth, and faith in the new china.* Macmillan.

Pils, E. (2014). *China's human rights lawyers: advocacy and resistance.* Routledge.

PRC State Council General Office. (2016). Guowuyuan bangongting guanyu zai zhengwu gongkai gongzuo zhong jin yibu zuohao zhengwu yuqing huiying de tongzhi. Retrieved from https://goo.gl/983hh8

Repnikova, M. (2017). *Media politics in china: Improvising power under authoritarianism.* Cambridge University Press.

Repnikova, M. (2018, November). *China's 'responsive' authoritarianism.* Retrieved from https://www.washingtonpost.com/news/theworldpost/wp/2018/11/27/china-authoritarian/?noredirect=on&utm_term=.749732649430

Roberts, M. E. (2018). *Censored: Distraction and diversion inside chinas great firewall.* Princeton University Press.

Schurmann, H. F. (1968). *Ideology and organization in communist china. enlarged.* University of California Press.

Stockmann, D. (2013). *Media commercialization and authoritarian rule in china.* Cambridge University Press.

Su, Y., & He, X. (2010). Street as courtroom: state accommodation of labor protest in south china. *Law & Society Review*, *44*(1), 157–184.

Tager, J., Glenn Bass, K., & Lopez, S. (2017). *Forbidden feeds: Government controls on social media in china.* Pen America. Retrieved from https://pen.org/wp-content/uploads/2018/06/PEN-America_Forbidden-Feeds-report-6.6.18.pdf/

Tang, W. (2016). *Populist authoritarianism: Chinese political culture and regime sustainability.* Oxford University Press.

Tanner, M. S. (2004). China rethinks unrest. *The Washington Quarterly*, *27*(3), 137–156.

Tarrow, S. (2005). *The new transnational activism.* Cambridge university press.

Wang, Y. (2016a, March). *The business of censorship: Documents show how weibo filters sensitive news in china.* Retrieved from https://web.archive.org/web/20161003182432/https://cpj.org/blog/2016/03/the-business-of-censorship-documents-show-how-weib.php

Wang, Y. (2016b, March). *Read and delete: How weibo's censors tackle dissent and free speech.* Retrieved from https://web.archive.org/web/20170515022534/https://cpj.org/blog/2016/03/read-and-delete-how-weibos-censors-tackle-dissent-.php

Wang, Y., & Minzner, C. (2015). The rise of the chinese security state. *The China Quarterly*, *222*, 339–359.

Wei, Y., Tian, X., Wang, H., Nie, G., Du, J., Feng, W., ... Deng, Y. (2015). *Gongan jiguan yuqing fenxi yu yulun yindao.* China Legal Publishing House.

Wintrobe, R. (2000). *The political economy of dictatorship.* Cambridge University Press.

Yang, Y., Yang, Y., & Fei, S. (2017, July). *China seeks glimpse of citizens' future with crime-predicting ai.* Retrieved from https://www.ft.com/content/5ec7093c-6e06-11e7-b9c7-15af748b60d0

Zeng, S. (2015). *Wangluo yuqing yingdui jiqiao.* Guangdong People's Publishing House.

Zhou, B. (2011). *Weibo wenzheng yu yuqing yingdui.* People's Publishing House.

Zhou, T. (2012). *Buduan kuochong shehui yulun de gongtong xinxi jichu.* Study Times.

Zou, C. L., Hongqiang, & Su, G. (2015). *Lingdao ganbu: Wangluo yuqing gongzuo zhinan work guide for public opinion.* People's Daily Press.