



Fear, Friction, and Flooding: Methods of Online Information Control

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

Citation	Roberts, Margaret Earling. 2014. Fear, Friction, and Flooding: Methods of Online Information Control. Doctoral dissertation, Harvard University.
Citable link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:12274299
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, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

©2014 – Margaret Earling Roberts
All rights reserved.

Fear, Friction, and Flooding: Methods of Online Information Control

Abstract

Many scholars have speculated that censorship efforts will be ineffective in the information age, where the possibility of accessing incriminating information about almost any political entity will benefit the masses at the expense of the powerful. Others have speculated that while information can now move instantly across borders, autocrats can still use fear and intimidation to encourage citizens to keep quiet. This manuscript demonstrates that the deluge of information in fact still benefits those in power by observing that the degree of accessibility of information is still determined by organized groups and governments. Even though most information is possible to access, as normal citizens get lost in the cacophony of information available to them, their consumption of information is highly influenced by the costs of obtaining it. Much information is either disaggregated online or somewhat inaccessible, and organized groups, with resources and incentives to control this information, use *information flooding* and *information friction* as methods of controlling the cost of information for consumers. I demonstrate in China that fear is not the primary deterrent for the spread of information; instead, there are massively different political implications of having certain information completely free and easy to obtain as compared to being available, but slightly more difficult to access.

Contents

1	Introduction and Motivating Examples	1
1.1	Information Friction: The Case of Air Pollution in China	6
1.2	Information Flooding: The Puzzle of Political Campaigns in China	9
2	Methods of Information Control	13
2.1	Previous Theories of Information Control	13
2.1.1	Traditional Media and Information Control	14
2.1.2	New Media and Information Control	15
2.2	Theory of Information Control	18
2.2.1	The Players: The State, Organized Interests, the Media and the Consumers	19
2.2.2	Degrees of Information Accessibility and Aggregation	21
2.2.3	Players and Information Control	26
2.2.4	Methods of Information Control	30
2.2.5	Preferential Access to Information Undermines “Free” Information	34
2.2.6	Competition Among Groups Creates “Free” Information	35
2.3	Structure of This Manuscript	36
3	Fear and the Paradox of Visible Censorship	38
3.1	Data	42
3.2	Limitations	42
3.3	Test One: Do Bloggers Change Topics After Censorship?	43
3.3.1	Data and Methods for Testing Blogger Self-Censorship	44
3.3.2	The Structural Topic Model	45
3.3.3	Results	46
3.3.4	Implications for the Spread of Information	55
3.4	Test Two: An Experimental Study of Consumers of Blogposts	55
3.4.1	Design	57
3.4.2	Results	58
3.5	Conclusion	61

4	Information Friction and Protests in China	62
4.1	Content Filtering Creates Information Friction in China	64
4.1.1	Internal Government Conflict Reduces the Efficacy of Censorship	69
4.2	Three Examples of Protests in China With Varying Levels of Censorship	70
4.2.1	Monolithic Censorship: Self-Immolations in Tibet	71
4.2.2	Local Conflict: Environmental Protests Against Local Projects	72
4.2.3	Larger Ideological Conflict: Anti-Japanese Protests	74
4.3	Content Filtering Slows the Spread of Information: Evidence	75
4.3.1	More Discussion of Self-Immolations on the Weekends When Censorship is Lower	76
4.3.2	Environmental Protest: Local Censorship Influences Local Discussion . . .	81
4.3.3	Anti-Japan: Websites with Higher Censorship Have Less Discussion	85
4.4	Conclusion	90
5	Newspaper Coordination and Its Internet Impact	91
5.1	Information Flooding and Propaganda	92
5.1.1	Intra-Party Conflict and Coordination	95
5.2	Information Flooding in China	96
5.3	The Influence of Flooding on the Spread of Information	97
5.3.1	Data	98
5.3.2	Identification of Propaganda Themes	99
5.3.3	Selecting Representative Articles and Measuring Coordination	102
5.3.4	Traditional Propaganda Articles are Most Coordinated	103
5.3.5	How Does Coordination Influence the Spread of Information?	105
6	Conclusion: Implications for the Information Age	112
6.1	Implications for Autocracies and Democracies	114
6.2	Implications for Policy	116

List of Figures

1.1	2011-2012: U.S. data for Beijing air pollution (purple) versus Chinese data for Beijing air pollution (red)	7
1.2	Time plots of propaganda topics in China	11
2.1	Information accessibility and information aggregation	23
3.1	Top topics across all of the blogposts within the study. The words presented are the most frequent words within each topic	48
3.2	Topical clusters. A line between two clusters indicates that the topics are correlated. Colors indicate the type of cluster.	49
3.3	Topic incidence for bloggers whose censored post was political.	50
3.4	Shift in political topics, before and after censorship.	51
3.5	Shift in life topics, before and after censorship.	52
3.6	Relationship between sentiment and posts before and after censorship.	54
3.7	Error page indicating that a post has been removed.	56
3.8	Treatment effect of censorship on the probability on next clicking on a post of the same topic.	59
3.9	Treatment effect of censorship on the number of posts selected on the censored topic after treatment.	60
4.1	Length of bursts for 120 self-immolation events between 2011 and 2013	77
4.2	Weekend self-immolations have longer bursts than weekday self-immolations, negative binomial model.	80
4.3	Geography of protest events for Qidong and Shifang environmental protests. Censorship begins with high variation across provinces but ends with low variation. White provinces indicate there is not enough data to provide censorship information.	82
4.4	Marginal effect of censorship on the number of posts in the following day, across cities in Qidong and Shifang environmental protests.	84
4.5	Websites with higher levels of censorship also have fewer posts, anti-Japan protests.	87
4.6	Censorship profile during anti-japanese protests over two websites. Red line indicates censored post, black line indicates all posts.	89

5.1	Propaganda themes identified with the Structural Topic Model.	101
5.2	Average level of coordination for most frequent topics.	104
5.3	Relationship between average coordination and number of total Google search results, by topic.	106
5.4	Relationship between average coordination and number of Google Search results on the site sina.com.cn, by topic,	107
5.5	Relationship between average coordination and number of Google search results on the site blogspot.com, by topic.	108
5.6	Expected increase in search results if coordination increases from 10 to 11 newspapers.	110

Acknowledgments

This manuscript would not have been possible without the support of institutions, colleagues, friends and family. I thank the Institute for Quantitative Social Science and the Weatherhead Center for International Affairs at Harvard University for financial support for my research in China. I thank the Ashford family for their generous funding and personal support throughout my time at Harvard through the Ashford Fellowship. I am indebted to Crimson Hexagon, which was generous with their data and support throughout my time in graduate school.

I owe a great debt to my many colleagues in graduate school whose ideas, criticisms, and friendship have enriched my time at Harvard and vastly improved my work. My conversations with Jennifer Pan and Brandon Stewart throughout graduate school have greatly affected my work. My dissertation reading group – Shelby Grossman, Jennifer Pan, Chiara Superti, and Amanda Pinkston – read many versions of my work and provided invaluable feedback. The residents of office K351, including John Marshall and Leslie Finger provided great conversation and humor through the last few years of writing.

I offer thanks to those people who supported me during my fieldwork in China. In particular, Kai Lukoff and Maoliang Ye provided incredible support and valuable conversations during my time in Beijing. Peter Volberding and Jeff Javed were the best fieldwork roommates anyone could ask for. Thanks to Frances Chen, Yichen Guan, Adam Jin, Heather Liu, Junjie Liu, Fei Meng, Hongmiao Wang, and LuShuang Xu for their assistance this past summer. Thanks also to those who helped me who prefer to remain anonymous.

Many others provided valuable feedback throughout the process. Other than those I have already listed, Any Catalinac, Joan Cho, Jeff Colgan, James Fowler, Judy Goldstein, Sheena Greitens, Justin Grimmer, Andy Hall, Jingkai He, Nancy Hearst, Haifeng Huang, Kyle Jaros, Ashley Jester, Iain Johnston, Aaron Kaufman, Jeehye Kim, Chris Lucas, Gwyneth McClendon, James Ashley Morrison, Rich Nielsen, Maggie Peters, Jake Shapiro, Susan Shirk, Daniel Smith, Jim Snyder, Arthur Spirling, David Steinberg, Alex Storer, Felicity Vabulas, Jessica Weiss, and Vanessa Williamson. For those who I have forgotten to list, it is not out of lack of appreciation but rather out of my own forgetfulness.

My friends Kate Hadley and Megan Hansen provided support throughout graduate school, and always reminded me to keep my work relevant. Dunster House, in particular Roger and Ann Porter, Carlos Diaz, and the amazing tutors and students there who became close friends over the past three years, were always interested in engaging in conversations about my dissertation, and I found those conversations to be very helpful.

My committee – Gary King, Jeff Frieden, Elizabeth Perry, and Dustin Tingley – have been everything a dissertation committee could possibly be. They have provided me with incredible feedback and support throughout my time at Harvard. They have been outstanding examples of scholars and teachers, and I have learned a tremendous amount from them.

My family has been the biggest source of support throughout my time in graduate school. First

and foremost, thanks to my husband David for your unconditional love, never-ending support and patience. You are incredible. Thanks to my siblings for everything they've done, in particular Emily for the amazing line-edits of my dissertation, Heather for Songza playlists that kept me editing, and Allen for jokes over G-chat and gourmet meals. Thanks to my first teachers and mentors, my parents Don and Barbara, for 28 years and counting of thought-provoking conversation. I dedicate this dissertation to them.

Chapter 1

Introduction and Motivating Examples

On January 12, 2010, Google disclosed to the public that it had been hacked by Chinese sources. In response, Google threatened to stop obeying laws required of foreign businesses operating in China and stopped censoring content returned within its search algorithm. The resulting conflict between Google and the Chinese government came to a head in March of 2010, when Google began redirecting its mainland users to google.com.hk, Google's unfiltered Hong Kong search site.¹

Reacting to Google's refusal to filter content, the Chinese government began blocking the unfiltered Hong Kong site. In addition, throughout the mainland, the government throttled access to Google services, allowing access to Gmail and Google's social media site Google Plus for only about 25% of requests (Milward June 30, 2011). China Unicom announced that it had removed Google search from its search platform, further restricting access to Google for Chinese users. Mainland users could access Google, but doing so required more work and more patience. For the first time, traffic to the search giant crept along slowly.

¹In a talk at Google headquarters, Google founder Sergey Brin claimed that the motivation behind Google's actions in China were based on his own experiences with information control growing up in the Soviet Union, although others claimed that Google was already being outcompeted by Chinese search engine Baidu and therefore had very little to lose from disobeying Chinese laws (Gustin March 25, 2010).

Google spends millions of dollars a year making search faster because research has found that faster search means more users. Google market research shows that slowing search by 400 milliseconds creates a .44 percent drop in search volume (Hoelzle January 2012). Other research demonstrates that slow internet is aggravating enough that 80% of internet users will get frustrated and leave Google's Youtube if loading a video stalls. Consistent with this research, when China began slowing Google's services in China, Google lost a large portion of the Chinese market. Since 2010, Google's market share in China has declined precipitously. In 2010, Google accounted for almost a quarter of the search engine market in China (Hille October 19, 2010), but its share had dropped to only 3% in 2013 (Slegg July 9, 2013).

In the summer of 2010, I was in China doing fieldwork trying to access my own e-mail on Google. The majority of young Chinese internet users I interviewed who had used Gmail before January 2010 had switched to Chinese service providers. Many were not aware that Google's slow speeds were government-created. A minority of users were aware that Google was being throttled by the government and used Google anyway, putting up with the persistent throttling *because* doing so was in direct defiance of the Chinese government. Since they knew Google's slow pace originated with the government, the act of protest was worth the cost of convenience.

While I was "jumping the Firewall" by accessing Gmail using a virtual private network (VPN), none of the students I talked to that summer had decided to continue to use Google by downloading a VPN. Indeed, a survey I conducted about internet use in China suggest that very few Chinese netizens jump the Firewall. Such software is relatively easy to find, but can be frustrating to use because it slightly slows down internet speed. Like throttling Google, the Firewall created small costs of access that continue to have significant implications for internet traffic in China.

In this manuscript, I demonstrate that small costs of access to information induced by censorship have a significant influence on the spread of information. "Information frictions" such

as throttled access to information, increased costs of access to information, or inconvenient data access can be almost as effective as complete censorship in its ability to keep information from consumers. States and other organized interests, who control infrastructure and who have incentives sufficient to spend resources disaggregating or disseminating information are able to determine the extent of friction for many types of information and thus control its spread. Whereas visible censorship often backfires to create “protesters”, such as the Chinese who were aware of throttling and yet continued to use Google, small changes in the costs of information are usually imperceptible, and therefore netizens may not know that their access to information is being manipulated.

I offer a new perspective on the methods of control of online information and their efficacy. Since the advent of the internet, many have discussed how this new medium would affect the relationship between the state and its citizens. Some argued that the internet would empower the masses, allowing them to spread information quickly and organize groups that could counter state power (Taubman 1998; Yang 2009). Others argued that the internet, like all other tools, would eventually be co-opted by the state, as the state could still completely censor information and implement laws that would incentivize citizens to self-censor online (Keohane and Nye Jr 1998; Boyle 1997; Hughes and Wacker 2003). These scholars posit that the state could use its monopoly over physical power to control any tool, even one as dynamic as the internet.

The discussion about how the age of information has altered the power dynamic between state and society is incomplete. States are not the only groups that can manipulate information in the digital age, and fear is neither the primary nor most effective method through which this manipulation occurs. While some information is accessible, even more is disaggregated, inaccessible and costly to obtain. Groups including the state, branches of the state, and other interest groups that have concentrated interests in gathering and distributing information are those most likely to control whether this information is accessible. These groups have an advantage in influencing the

spread and impact of information in the age of the internet because they are willing to pay the costs of obtaining and packaging this information. Since these groups wield control over the information they gather, they have the ability to make information expensive or cheap for the media and for the public. Further, these small manipulations largely go unnoticed by the public. In a time where citizens have much more information to consume than time to consume it, small manipulations of the costs of information have significant effects on the spread of information and, in turn, on public opinion and collective action.

Unlike previous scholars, I propose that the state uses strategies similar to those of organized interests to promote its objectives online and in the media. The state, like other organized interests, controls the spread of information by manipulating the costs of information access. While others have argued that the state's monopoly on fear allows it to control the spread of information by intimidating producers and consumers of information, I find that such highly observable censorship and fear actually undermine the reputation of the state. Instead, largely unobservable censorship and propaganda implemented through small increases and decreases in the cost of information, which I call *information friction* and *information flooding*, are the primary mechanisms through which states with high levels of internet penetration can control the spread of information in the information age.

Depending on the type of information and political environment in which the group operates, information flooding and friction strategies will be easier or more difficult to implement. First, in cases where collecting information is very costly and can only be collected by particular groups and not the public, when information is *excludable*, flooding and friction are under the control of one or a small number of groups and will serve these groups' interests. For example, government officials are first privy to internal government secrets, or issues with product safety are first revealed and understood by the producing company. These groups then act as gatekeepers in the dissemination

of this information, making the costs of information easy to manipulate.

The second parameter that influences the success of friction and flooding strategies is the number and extent of competition between organized groups that are interested in collecting a particular type of information. Even if information is excludable, competition between two groups that gather the same information may create incentives for the information to be leaked and spread amongst the wider public. For example, competing bureaucracies both privy to particular government information might leak government secrets that cast a negative light on the other group. Competition, in this case, creates incentives for the dispersal of otherwise difficult-to-access information.

The fact that the state controls information in a similar way as organized interests, and may in some cases face information competition from organized interests, implies that information manipulation is not specific to authoritarian regimes with strict information laws. Political manipulation of the spread of information can occur in any society, by any group that has incentives to collect and control information. The availability of information is determined not by political regime type, but by the degree of competition among organized interests, the degree to which these groups have preferential access to information, and the incentives of these groups for manipulating access to this information. These factors vary significantly within countries and across issue areas, meaning that the prevalence of censorship is much more fine-grained than country-level investigations have led us to believe. It also implies that the most powerful form of censorship can be initiated by any organized group, using methods that are largely unobserved by the public. While the evidence within this manuscript is drawn primarily from China, future research will extend the same principles to organized groups and governments in other countries.

The manuscript is organized as follows: First, I present two vignettes to provide an intuition behind information friction and flooding in China. In Chapter 2, I explain my theory of information friction and flooding in detail. Chapter 3 provides empirical evidence from an experiment in

Beijing and large data analysis on the internet that fear is not the principal mechanism responsible for information control in China. Chapters 4 and 5 provide empirical evidence that information friction and flooding, respectively, are the main mechanisms behind information control in China. Chapter 6 draws implications for information control in other countries, with a focus on the United States, and concludes.

1.1 Information Friction: The Case of Air Pollution in China

Beijing is one of the most polluted cities in the world; sixteenth most polluted, according to the World Bank. Air pollution in China is thought to have caused 1.2 million premature deaths in 2010 alone, which accounted for 40% of premature deaths in the world that year (Wong April 1, 2013). However, data on pollution is largely kept from citizens in Beijing. The U.S. embassy takes air pollution measurements in Beijing hourly and publishes them on its Twitter feed, but this feed is blocked in China. Instead, China releases its own air pollution data, which are consistently lower than U.S. measurements. A comparison of Chinese and U.S. pollution data for Beijing is provided in Figure 1.1.

The fact that Beijing censors data on air pollution might be puzzling to those who contend that the internet should spread information quickly, and that it would be impossible to keep online information away from citizens. Although Beijing censors air pollution data by blocking it with the Firewall, air pollution data is not impossible to obtain. In fact, accessing this data is as simple as installing a VPN. All 800,000 Beijing internet users could quite easily connect to accurate, hourly pollution data.

Further, air pollution is easy for citizens to observe daily. While the exact measures and implications of pollution numbers are censored, extremely polluted days in Beijing are obvious since on those days visibility is low in Beijing, and “blue-sky” days have lower levels of pollution. It is

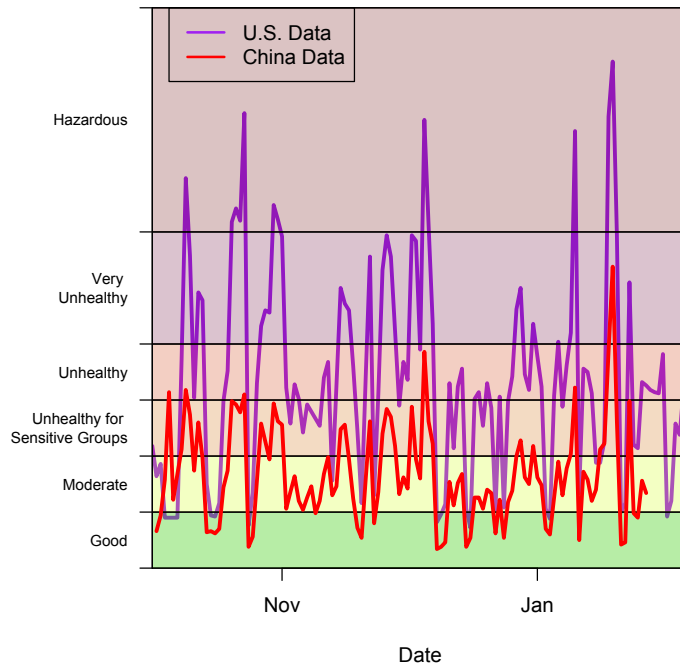


Figure 1.1: 2011-2012: U.S. data for Beijing air pollution (purple) versus Chinese data for Beijing air pollution (red)

apparent that some days, weeks, and months are less polluted than others. In addition, pollution monitors are not terribly expensive, and can be used by normal citizens in China. Citizens could technically measure their own pollution data, instead of relying on the government to collect it for them.

However, these small costs of jumping the Great Firewall or collecting data individually had a significant influence on people's perception of pollution in Beijing, at least for a period of time. Very few people in China download VPNs, and, therefore, very few people know about the discrepancies between Beijing-reported air pollution and the real air pollution numbers. In my conversations with Beijing residents in 2010, some referred to heavy pollution days as "foggy".

Furthermore, China's censorship of air quality data obfuscates the health danger that Beijing's air pollution poses. Because Beijing and the embassy's versions of the data resemble each other, residents might interpret Beijing's data as simply on a "different scale" than the U.S. version. However, Beijing's data never reaches the "Hazardous" level, U.S. measures do frequently. Censorship and Beijing's own collection of data allows it to control access to the data as well as its interpretation. By making unfavorable data slightly more difficult to access and promoting more favorable data, China shapes public perception of Beijing air pollution.

When do such small frictions cease to be effective in slowing the spread of information? In this manuscript I argue that another organized group, with a similar incentive to aggregate and distribute data, has the ability to thwart the power of information friction. The cost of air pollution data in Beijing decreased dramatically in January 2012, when an organization called Green Beagle, whose stated purpose is to collect and distribute information about pollution and its health consequences, began posting the U.S.-produced Twitter information on its own, unblocked website in China. The organization encouraged Beijing residents to buy their own pollution monitors and post their own information on the web (Watt December 8, 2011). Suddenly, with lower-cost pollution data on hand, Chinese residents began with demand better information from the government. In the same month, more than 1500 petitions were filed to the government requesting the release of better information ("Chinese public want to see changes in monitoring pollution." December 9, 2011a). Caught in the public eye, in January 2012 Beijing announced it would release such data in 2016, and also made unexpected international concessions on global warming.

1.2 Information Flooding: The Puzzle of Political Campaigns in China

Since Deng Xiaoping, each new leader of China has formed his own political slogan that identifies his administration's outlook on governance. Deng's slogan, "socialism with Chinese characteristics," implied that China could remain a socialist country while developing a market economy. Jiang Zemin, who followed Deng, introduced the "Three Represents" in the year 2000, indicating that the Chinese Communist Party should represent three fundamental principles: developing productive forces in the economy, progressing Chinese culture, and representing the interests of the masses. Hu Jintao introduced "scientific development outlook" when he first entered office, focusing on balanced development among social classes and environmentally sustainable development. This was followed by "harmonious society" in 2005, two years after he had entered office, to indicate that the Party should focus on harmony between conflicting groups in society. Most recently, over the last year, Xi Jinping has introduced "chasing the Chinese dream," which focuses on rejuvenating the Chinese nation.

Communist party slogans not only communicate to the public the outlook and ideology of the current leader, but also are a method by which leaders assert symbolic power over the Party and population (Perry 2013; Huang and Li 2013; Brady 2009). These slogans are plastered throughout China, on posters, within schools, and in newspapers. Recently, for example, schools have had Chinese dream competitions, workplaces have selected "model dreamers," and Communist Party singer Chen Sisi has performed the song "Chinese dream" hundreds of times on television and before large audiences ("Chasing the Chinese dream." May 4, 2013). In Figure 1.2, my own analysis of Chinese newspapers (which I will describe in more detail in Chapter 5), shows the progression between the topics of scientific development, harmonious society, and the Chinese

dream within provincial newspapers over Hu and now Xi's time in power. These newspapers are highly coordinated in their printing of the slogans, quickly picking up and quickly dropping slogans based on the leader in power.

Because the Chinese government uses the power of its organization to flood the information environment with the slogans, these slogans are perhaps public information with the lowest cost for the Chinese people. Indeed, it would take substantial effort to avoid the slogans. The slogans permeate not only the rhetoric of the CCP, but also the language of normal citizens. A Google search indicates that there are around 71,000 mentions of "harmonious society" on blogs on sina.com.cn, 106,000 results for "scientific development outlook," and 56,000 results for "chasing the Chinese dream." International blogs repeat the slogans as well; indicating that the influence of the slogans spreads even beyond the Great Firewall. A similar search on blogspot.com (which is blocked in China) recovers thousands of search results.

Why do so many people incorporate the slogans into their language? It is obvious that these sayings are propaganda slogans, and their informational content is marginal at best. While citizens would undoubtedly consume these slogans, why should they repeat them? It is unclear how these slogans will translate into specific policies, or what any individual should do to respond to them. The ambiguity of their real world implications as well as the obviousness that they originate from a government source, should undermine the slogans' influence on citizens within China.

In this manuscript, I demonstrate that information flooding allows information that is very easy to consume to seep into the subconscious of individual citizens. Information that is distributed in a coordinated way by an organized group or government is very difficult to avoid and remains on the minds of citizens. These citizens then repeat this information in their own writing, creating a multiplier effect that allows the slogan to pervade the blogosphere and the language of citizens within China. By dominating the conversation, the flooding also creates less room for less desirable

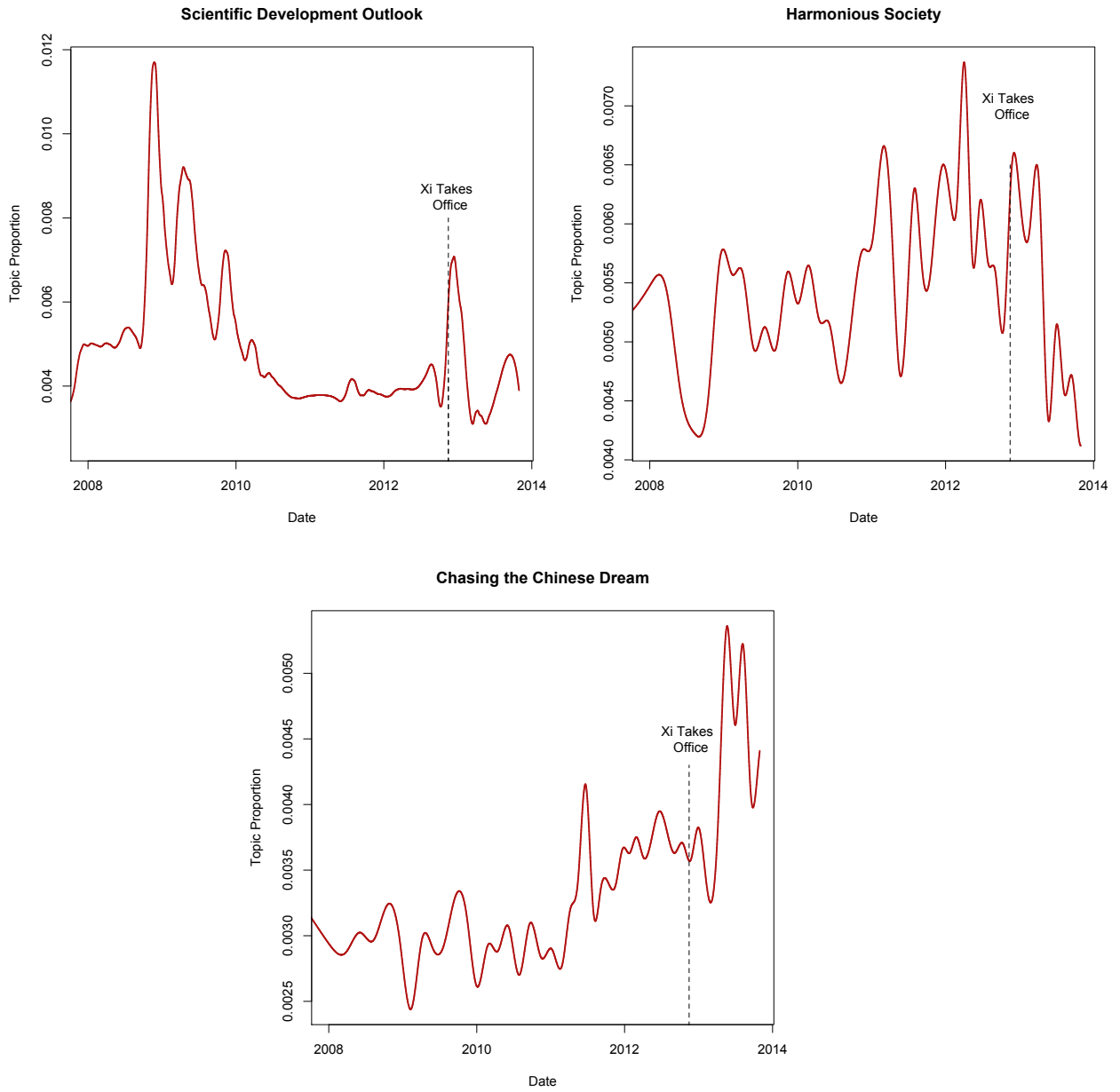


Figure 1.2: Time plots of propaganda topics in China

topics, such as air pollution discussed in the previous section. Contrary to popular notions that the internet allows citizens to access information that is most important to them, these examples illustrate the fact that information control may yet still be possible in the digital age.

Chapter 2

Methods of Information Control

This chapter delineates a theory of information control based on small changes in costs of access to information. I maintain that organized groups and the state, not the media or the average consumer, are most likely to control the costs of information access that have important effects on the spread of information. To begin, I overview previous perspectives on information control, then introduce the main players in the media environment and the logic behind the theory. Last, I describe situations in which information control is more and less effective.

2.1 Previous Theories of Information Control

In this section, I outline some of the previous literature on information control. I begin with studies of traditional media, such as newspapers and television. I then discuss others' expectations that new media will change both the ability of the state to control information and the methods by which this control is exerted.

2.1.1 Traditional Media and Information Control

Previous literature on information control in traditional media sources has attributed the media's ability to provide "free" information to the ownership and market structure of the media. Siebert (1956) defines different philosophical bases for media. The libertarian view, which provides the basis for the American press, is founded on the idea of media independent from the government. In contrast, in the communist or authoritarian view of the press, the press acts as an extension of the government, communicating messages to its citizens.

Djankov et al. (2001) measure government ownership of the media in 97 countries and find that in the vast majority of these countries, governments own a substantial portion of the press. They find that government ownership and control of the press is negatively correlated with political and economic freedoms in these countries. Others have gone one step farther to argue that greater competition among media increases the quality of information provided to the public. These scholars find that competition between newspapers decreases bias (Qin, Wu and Strömberg 2012; Galvis, Snyder Jr and Song 2012) and makes it more difficult for governments to pay off or intimidate media sources as the number of media sources increases (Besley and Prat 2006). Others find the opposite: that increased competition among newspapers can actually exacerbate bias, as it reduces the quality of reporting and shifts the focus of news toward entertainment, or the likely views of consumers, and away from neutral reporting on important political issues (Mullainathan and Shleifer 2005; Cagé 2013).

Over the past twenty years in China, the newspaper market has gone through a significant transformation. Beginning in the 1990s, the Chinese newspaper market underwent what Stockmann (2012) terms *media marketization*: the deregulation, commercialization, and privatization of newspapers. This transformation incentivized media to respond to market forces, as they no longer could rely on state funds. While some have argued that commercialization has increased

local government responsiveness to media and increased newspapers' ability to investigate and uncover local scandals (Tong and Sparks 2009; Tong 2011), the intricate propaganda infrastructure of the CCP still controls sensitive political messages among newspapers. Stockmann (2012) finds that media commercialization may actually solidify central control and maintain the propaganda infrastructure by providing credibility to local papers. Others have argued that commercialization of the media buttresses the power of the central government by keeping local governments in check (Lorentzen 2013). While media marketization has had some benefits for information freedom in China, it has not solved the problem of bias within the Chinese media.

2.1.2 New Media and Information Control

Given the speculation that increased competition might decrease media bias, the introduction of new media created much excitement among scholars and policymakers. In 2000, President Bill Clinton announced that trying to control the internet is equivalent to "trying to nail jello to a wall." Many assumed that new media would make the information environment more free because it was impossible for a government to "own" new media and the proliferation of media sources caused by the internet would create increased competition (Gentzkow and Shapiro 2008). The internet was credited with promoting democracy in previously authoritarian countries such as Indonesia, Malaysia, and Serbia (Ferdinand 2000). It is also cited as a critical ingredient in the rise and spread of protests in the Arab Spring (Ada et al. 2012; Bellin 2012).

In China, some predict that the internet could have serious destabilizing effects on the CCP's efforts to retain power. Although the CCP exerts some control over the public's access to foreign websites, the porous nature of the internet means that such attempts cannot be wholly successful (Harwit and Clark 2001). Latham (2007) distinguishes between "orderly" and "disorderly" media, arguing that "disorderly" media, such as the internet and text messaging, pose a threat to the

traditional propaganda/state-controlled media in China and, by extension, to the government itself.

The Chinese government has invested in and encouraged the development of internet companies in China because it sees it as a route to economic development, but scholars have argued that this development may also undermine the government's propaganda efforts (Taubman 1998). Internet companies have incentives to avoid or delay the censorship of information in order to pander to users; the market benefits companies that ignore censorship directives. While initially only wealthy, young people with state buy-in used the internet, Damm (2007) suggests that as the internet expands to more marginalized groups, more people could express their discontent and possibly threaten the regime.

Even those who acknowledge that the internet may not bring democracy to China believe it has given citizens more say in policymaking. Yang (2009, 2014) claims that the internet fosters online activism and encourages the formation of digital civil society. Shirk (2011) argues that the internet provides a check on local governments, alerting the central government to local problems before local governments can cover them up. Zheng (2007a) contends that unlike some other types of reforms in China, the internet has benefited the poor more than the wealthy.

On the other hand, some maintain that because the state has a monopoly on force, it can control any technology that threatens its hold on power. Keohane and Nye Jr (1998) point out that even though the information revolution changes technology, it does not change the contours of military power and therefore will not significantly change the state-society power structure. Lessig (1999) observes that since the state can control the underlying code and the laws that shape the structure of the internet, it will control how the internet influences society. Harwit and Clark (2001) contend that because the state cannot technologically control the spread of information, self-censorship by fear will be the main way through which the government controls content on the internet.

Hughes and Wacker (2003) claim that under authoritarian governments, a history of censorship

and intimidation will make society less likely to benefit from the internet. Unlike democratic societies in which citizens trust the state to allow freedom of expression online, the “lack of trust capital” and uncertainty about what the state can control on the internet will prevent netizens from using the internet to its full capacity. The state’s monopoly on force will outweigh the potential for the masses to take power using the internet.

In the literature on social media in democracies, most scholars maintain that the internet has increased public access to information and expanded public participation in policymaking (Dahlberg and Siapera 2007; Trippi 2005; Tolbert and McNeal 2003). Widespread access to the internet means more people have access to more information and policymakers must respond in realtime to public opinion. However, the way in which new media influence governance within democracies is still subject to much discussion. While some authors suggest that this new type of “e-democracy” is a more pure, more Athenian version of democracy (Coleman and Gotze 2001), others argue that empowering the masses undermines elite influence over public policymaking and leads therefore to a deterioration of governance (Bellamy and Raab 1999). Still others maintain that new media is a distraction to citizens, affording a false sense of participation in policymaking, when other types of political participation such as joining a civil society group or attending a city hall meeting would in fact be more meaningful. Whether the masses “should” exert such increased influence through new media is the point of most contention in the academic literature on the effect of the information age on the relationship of citizens to the democratic state.

Like these scholars, this manuscript addresses the question of who controls the spread of information in the digital age. However, I examine the collection and costs of information as well as its dispersal. Who collects and distributes information, and how does this influence its spread? Instead of thinking about information control as a conflict between state and society, I seek to better understand which actors have interests in gathering and distributing information and which

actors will only consume the most accessible information. Organized interests with money and incentives to produce and distribute information continue to wield the most power over the public and the media, both print and online, both old and new. At times, the state is this organized group, spending enormous amounts of money to produce and distribute information that will shine a favorable light on its performance. At other times, factions within the state, bureaucracies, trade sectors, non-governmental organizations (NGOs) and companies are the critical actors, hiring their own experts to produce and distribute information and withholding other types of information that would cast them in an unfavorable light. I turn to this theory of manipulation in the information age in the next section.

2.2 Theory of Information Control

The central thesis of this manuscript is that small costs of access to information have a disproportionate influence on the spread of information and that the state and organized groups are the primary entities that influence these costs. In this section, I introduce the main players within this theory: the state, organized interests, the media and consumers. I delineate two different ways through which information can be costly: information can be made more *inaccessible* in medium, or more difficult to access through the traditional or new media, and information can be *disaggregated*, or exist in many different locations and therefore costly to collect. I claim that consumers will be more likely to read accessible and aggregated information because this type of information is low cost. The state and organized interests will primarily influence the costs of information by influencing its degree of accessibility and aggregation. I define *information friction* as the act of making information more costly, by disaggregating it or decreasing its accessibility, and *information flooding* as the act of making information cheaper, by making it more accessible and aggregated. I hypothesize that information friction and flooding will be the two primary and most

effective strategies of online information control, and that even small changes in accessibility and aggregation will have significant effects on the spread of information.

2.2.1 The Players: The State, Organized Interests, the Media and the Consumers

1. The State and Organized Interests

I categorize both the state and other organized interests as groups or organizations which have concentrated interests in particular policies or outcomes. These groups are small enough or have an organization strong enough that they are able to overcome the collective action problem to effectively pursue a shared objective (Olson 2009). They can be distinguished from larger diffuse groups within the public, which cannot overcome the free-riding problem to cooperate to achieve their common goals.

I address both states and organized interests within this category because at times the state is not a monolithic entity, but is divided into groups that could be organized interests or semi-autonomous institutions. Democracies can be divided into political parties, each with their own objectives and interest in their public image. Authoritarian governments may at times be broken into factions, where each faction has different public objectives (Nathan 1973; Shih 2008*b*). Bureaucracies within states may have divergent goals and objectives, and at times these entities may act as organized interests that oppose the objectives of other entities within the state (Etzioni-Halevy 2013). While at times the state may act on its own, more commonly the state is segmented into groups with conflicting interests (Allison and Halperin 1972). Further, groups within the state will act similarly to organized interests outside the state, such as companies, interest groups, or civil society organizations, and therefore for my purposes can be considered as part of the same category.

My theory only applies to groups that need public support for the furtherment of their objectives. For example, the state needs public support for its own survival. Political groups seek electoral support for a particular political leader or candidate. If the group is a company, public support could mean consumers buying the company's products, or investors buying the company's stock. If the group is an non-governmental organization (NGO), it could be donations or volunteers.

There are groups that do not need public support to achieve their common goals. Groups that get together to support each other, exercise together, or gain entertainment from each other obtain their objectives within the group, without need for outside support. However, groups that do not need public support can very quickly become politicized when their environment changes; for example, when the group is outlawed, or when a member is negatively affected by a policy. Therefore, this theory could potentially apply to almost all groups that have the ability to overcome the collective action problem.

2. The Independent Media

The next player in this theory is the independent media, which I define as organizations that generate revenue from writing and distributing news and whose employees do so as full-time work. The media in this case include traditional print, television, or online news.

Since many organizations and individuals produce media, I distinguish between individual bloggers who write news as a hobby, group-run newsletters, and the independent media, the latter of which I am concerned with here. Under my definition, the media does not include bloggers who write for entertainment purposes, or on the side of their full-time work. Though these bloggers are important players in disseminating information, I consider these part-time or hobby bloggers primarily as consumers of media, which I describe below.

In addition, I consider press that is captured by the state or is the spokesperson for an organized interest to be part of the state and organized interests, and not part of the independent media. In these cases, any information generated by the captured media is generated on behalf of the state or organized interest, and funding for this media does not necessarily come from generating news itself.

3. The Consumers

Consumers of media are the part of the public who read and share news in their leisure time, but who do not earn their living from creating or distributing news. These consumers are the primary audience of organized groups and the state for seeking public support. Since a consumer's employment is not to read or share news, she does so under severe time constraints and may not be very well informed. Consumers tend to rely on the independent media, the state, and other groups to determine what information is newsworthy. However, these consumers may themselves influence others within the public, as they may share news or information they read with their friends and family through social media or other mediums. As I explain later in this chapter, the news that consumers of media consume can have a multiplier effect on other consumers, and sometimes on the independent media itself.

2.2.2 Degrees of Information Accessibility and Aggregation

While it is often assumed that in the age of new media, all information is accessible, this section will dispute that view, delineating different degrees of information accessibility, from accessible information to inaccessible, and different degrees of information aggregation, from aggregated information to disaggregated information. Information accessibility and aggregation both lie on a continuum, and therefore there can be degrees of information accessibility and degrees of aggre-

gation. As I show in examples below, I distinguish between aggregation and accessibility, so that information can be aggregated, but relatively inaccessible, or very accessible, but disaggregated. Figure 2.1 summarizes the spectrum of aggregation and accessibility.

1. Accessible Information

Accessible information is information that, in medium, is readily available to the public. This information can be accessed at low cost and takes little effort to consume. The most extreme example of accessible information in traditional media are stories on the front page of all newspapers, with automatic subscriptions given to consumers. Online, this is information e-mailed directly to consumers or on the front page of websites that they frequently visit. It could be stories that air on televisions within airports, shopping malls, or other places consumers frequent. It could be information on primetime news, when most viewers will be watching. The most accessible form of information is information consumers may not even look for or try to consume, but end up consuming anyway.

2. Inaccessible Information

On the other extreme on the accessibility continuum, inaccessible information is information that in medium is very difficult to access. Consumers will not find this information in sources that they typically frequent, such as newspapers or the web. Even if it has been collected and aggregated, it may be extremely costly for the public to access because the groups that have collected it refuse to release it in an accessible medium. Completely inaccessible information, rare in the information age, is information that can be completely kept secret.

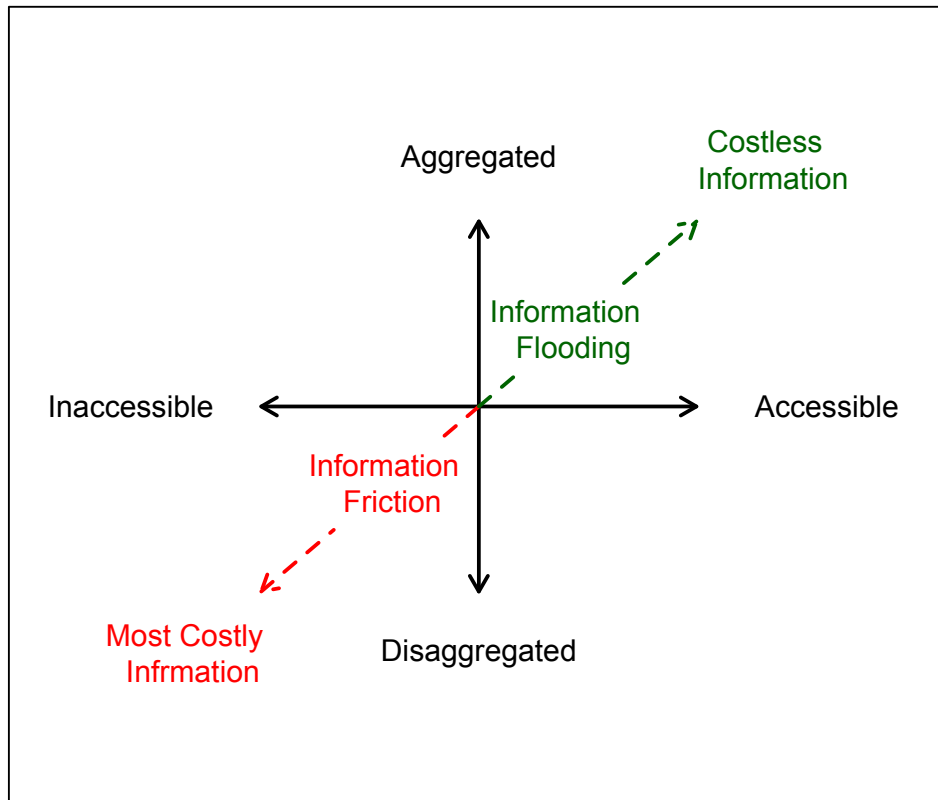


Figure 2.1: *Information accessibility and information aggregation*

3. Aggregated Information

Aggregated information is information where most of the facts necessary to convey the story or perspective to the public have been collected. Aggregation of information can occur through using surveys to collect information from a variety of people, and summarize those opinions, or by collecting data or accounts of an event from a variety of different sources and summarizing these accounts in one place. The aggregation of information will not necessarily be balanced, and what information is aggregated will depend on which group or actor collects the information. For example, a gun rights group may collect information about gun deaths, but not about the number of lives saved with guns. However, aggregation in any form makes information that used to be dispersed more available to the public, and therefore is similar to accessibility in that it decreases the costs of consuming that information.

4. Disaggregated Information

Disaggregated information is information that exists in many different places or with different people, but has yet to be collected and synthesized. Public opinion is by nature disaggregated information, as individual opinions exist within each citizen and need to be collected before they are meaningful. Public opinion polls are typically the entities that aggregate this information. Each piece of disaggregated information may be very accessible, but consumers would need to expend significant effort to aggregate the information and see the whole picture.

Accessible, Aggregated Information An example of highly accessible and aggregated information is group advertising, or government propaganda. A billboard in Boston sponsored by a gun-prevention group provides an updated count of the number of children who have died from gun violence in the current year. The group has collected and summarized this information into

one number. In addition, the information is highly accessible: passersby will notice this number, whether they seek it out or not.

Similarly, propaganda about a government-sponsored event or anniversary can be both highly accessible and aggregated for the public. A report on a government parade in China, for example, might aggregate accounts from many different participants, and coverage of the event might be on every major news station in China. Regardless of whether people choose to seek out this information, they might consume news about the event subconsciously.

Inaccessible, Aggregated Information An example of inaccessible, but aggregated information is data collected within governments, but that the government refuses to release. For example, the CCP collects data on the number of protests in each province and use this data to evaluate the performance of local government officials (Chen 2012). This data has been collected, and therefore is aggregated. However, the CCP makes the aggregated data extremely difficult to access because the number of protests reflects badly on the government. While some people with inside connections may be able to access the data, doing so is incredibly difficult and requires sufficient investment of time that is prohibitive to most media, academics, and consumers.

Accessible, Disaggregated Information An example of accessible, but disaggregated information are counts of particular types of crime incidents in the United States. Both state and local agencies, such as police departments, collect this data and most are required by the state to make this data available to the public online or by request; therefore, agency-level data are relatively accessible. However, these data are not completely aggregated at the national level. While the FBI created the National Incident-Based Reporting System (NIBRS) in 1991 to aggregate the data, incident numbers are volunteered by agencies, meaning NIBRS is far from complete. While the individual information is quite accessible, aggregating is very costly.

Inaccessible, Disaggregated Information Inaccessible and disaggregated information is the most costly type of information. Each piece of this type of information is in medium difficult to access, and the information will be disaggregated over different people, and places. An example of inaccessible, disaggregated information is knowledge of the extent of lead poisoning across China. Lead poisoning is one of the major threats to public safety, in particular children's safety, in China (Shen et al. 1996). Due to the expansion of factories in China that produce lead acid batteries for motorcycles and electric bikes, thousands of people across China are thought to have dangerously high levels of lead in their blood. Good information is very difficult to collect, as it requires measuring and interviewing large numbers of people who are difficult to access, both because they are spread out geographically and because of government barriers, including local governments and hospitals that refuse to administer lead tests (LaFraniere June 15, 2011; Cohen and Amon 2011). Even if this information were collected, to make this information accessible experts would have to analyze the results, summarize the results, and find a way to put the information uncensored online. Public health researchers have made progress in collecting this information and making it available to the public, but this process has taken a decade, an indication in itself of the costs involved in collecting this information and making it available.

2.2.3 Players and Information Control

This section explains which actors are likely to consume and produce which type of information. The information each produces and consumes will depend on 1) the incentives they have in making such information available and 2) the time and money they have to collect and make the information accessible. When their interests outweigh the costs of providing the information, the players will collect and provide that information to the public.

The Consumer Will Primarily Consume Low Cost Information

The typical consumer will consume mostly accessible, aggregated information. As information becomes more disaggregated or more inaccessible, they will be less likely to consume it. Consumers rely on others to provide pre-packaged information in the form of news or online media (McCombs and Shaw 1972). They have limited time to aggregate information or find inaccessible information. Further, they often only have marginal interests in seeking out information. Consumers are only interested in consuming information as much as they are interested in forming opinions about policies. In democracies consumers may be interested as voters and in autocracies as participants in protests or in political gatherings. Many in both cases are only interested in news as a hobby or as a citizen.

Information today is often accessible to the public online, but consumers can only consume a subset of this accessible information. They will largely allow the media and other groups to determine which content is most valuable for them, and will consume only what is on the front page of their favorite newspaper, or social media aggregator. Investing time and resources into extensive research seeking out information about particular policies will most often not be worth it to these individuals.

Consumers often distribute news by sharing accessible information with their friends and family. As discussed more in detail below, consumers are likely to re-share information or news that they find valuable. This information is likely to be a subset of low cost information, and therefore is already in a medium that is easy to share, such as in a newspaper or online. The result of this is if groups and the media can get the attention of a subset of consumers, they are likely to be able to grab the attention of a subset of their friends and family also.

The Media Has Fewer Resources to Aggregate and Digitize Information

The independent media's role is to determine what content is newsworthy, then collect and aggregate that information for the consumer. As described in the previous section, because media is usually very accessible, the independent media sets the information agenda for the public (McCombs and Shaw 1972). In some cases, such as investigative reporting or in-depth analyses, the media will transform disaggregated information or even inaccessible information into accessible information for the public. When the media are independent, they provide an important public good by making high cost information low cost for consumers, and they will follow the demands of the consumer.

However, it is difficult for consumers of information, with the little time they have, to tell the difference between news that was *already* accessible and news that the media aggregated or *made* accessible. Further, consumers cannot observe what information the media is *not* reporting, since the consumer is unlikely to do her own research into unreported news. As a result, the media has incentives to reprint news that is already accessible to save on the high costs of investigative reporting and aggregation of information, without losing credibility from consumers.

With the advent of the internet, the number of media sources has proliferated and has increased competition for traditional media sources, reducing their ability to invest in investigative journalism. Greater supply of accessible information means consumers are less likely to pay a high price for information. The dramatic expansion of sources forces traditional news to adapt to the low cost, high volume model of the internet in order to compete (Nichols and McChesney 2009). The internet has competed away revenues and therefore the resources for high cost investigative reporting have become less and less available. In the United States, many smaller newspapers have been forced to close due to the loss of readership to online media sites, as well as competition for advertising dollars with online display ad formats and free online classifieds such as Craigslist.

While online news sites have led to a modest decline in newspapers – the number of daily American newspapers decreased from 1,611 in 1990 to 1,387 in 2009 – it has also meant a dramatic reduction in profit margins. Full-time professional staff at newspapers dropped 27% between 1989 and 2010 (Edmonds et al. 2011). In particular, resources for investigative reporting have been cut, and newspapers are more and more reluctant to allow their short-staffed journalists covering daily news to invest in high-cost stories (Nichols and McChesney 2009).

Since consumers of media cannot necessarily differentiate high quality news from low quality news, media do not have significant incentives to spend resources making inaccessible information accessible and aggregating information. The media increasingly favor stories where information is fairly easy for it to obtain. The implication of the information age is that while more information is available to consumers, overall the degree of accessibility of information is determined by the cost of that information to producers of media, not necessarily its inherent value.

Organized Interests Are More Likely to Aggregate And Distribute Information

While the media are often not willing to pay the costs necessary to aggregate and make information accessible, organized interests and the state have concentrated interests in collecting and sometimes making accessible information about policies that relate to them. First, these groups are likely interested in aggregating information related to policies they are interested in for their own use. Information about these policies can help them lobby policymakers to support their cause (Potters and Van Winden 1992). For bureaucracies, gathering information related to their performance can help them justify their existence to their superiors. For the state, information can also help it conduct quality control on the activities of their subordinates. Second, these groups are most likely to use this information to educate and persuade the public. The state gathers information about its performance, not only for its own internal use, but also as a public relations tool. Since these

groups are all in need of public support to further their goals, aggregating information and making it accessible to the public is frequently key to their objectives.

Not only do organized groups have, in some cases, a greater interest aggregating information, they will sometimes also pay lower costs of collecting information since they have the expertise and initial lower costs of access to the information. For example, the groups can use their internal connections to gain access to interviews. The state itself has better access to employees within the state, because these employees are on their payrolls. The state also has inside knowledge about where to look for information, making information aggregation lower cost than for the public or media.

Last, organized groups have an advantage in disaggregating information that might reflect poorly on them. Since these groups are more aware of the sources needed to make information accessible, they can make these sources difficult to access by the media, or put restrictions on the types of information they collect and publish. For example, companies being sued for product safety infractions may seal these court cases to make this information less accessible to the media and the public. Groups can simply not collect information that might reflect badly on them, making it more difficult for others to aggregate. Organized groups themselves are often the primary sources for the media and public, and therefore their cooperation is imperative. Without it, it is much less likely that the information in question will ever be synthesized and accessible to the public.

2.2.4 Methods of Information Control

In the previous section, I argued that the state and organized interests often have the resources, interest and advantage in aggregating information. However, unlike independent media, these groups do not always have an incentive to make this information accessible. In some cases, the group uses its resources to hide this information from the public, making this information even

more disaggregated and inaccessible, moving information toward the bottom left of Figure 2.1. In other cases, the group will use its resources to aggregate the information so that it is extremely available to the public, moving information closer to the top right of Figure 2.1. In both cases, the group's ability to manipulate the aggregation and accessibility of information gives the group power over how costly that information is to the public and the media, and therefore determines what the public is most likely to consume.

For example, say a reporter or blogger wanted to write an article about traffic safety, or a consumer was interested in getting information about traffic safety. She would need to collect data on the incidence of traffic accidents, traffic citations, and number of cars on the road. She would ideally have information on the percentage of people wearing seat belts and emergency response times. She would want to know the location of crashes and road conditions and she would need access to police officers to be able to interview them for more detailed information.

The government already has incentives to collect data on traffic safety because it creates oversight of the police departments and could potentially point to areas needing road improvements. The government likely has been collecting this type of data for several years. Further, the government has low costs to collecting this information, since it can require police to record the data in a database. It is unlikely that our blogger, reporter, or consumer would collect the information herself, since it would be much more difficult to collect this information on her own by following police or ambulances around to record road safety accidents. Also, the blogger or reporter may only have a few days to write the story, so even if she were able to follow police officers around to collect data during that time, limiting her ability to personally collect the historical data.

If the government has collected this data, it might decide to make this information easily available to our reporter or consumer through state websites, or by contacting the police department. Or, it may be costly to obtain. It may technically be available, but the administrators in the police

department may be very unhelpful, or may require a series of fees or processing time to gain access to it. If it is sufficiently expensive, the reporter or consumer might give up and decide gathering the data is not worth it. Or, the reporter may only gain partial access to the information, but may be missing crucial information. In this case, because the government collects the data, the government itself, is the first to impose information access costs which could ultimately determine the likelihood that the information will become accessible to the public.

Information Flooding

When the government or organized group would like as many people as possible to see the information, the strategy I call *information flooding* the group aggregates information and makes it as accessible as possible to the public. The group might synthesize the information so that the media can easily reprint it at low cost and could try to persuade the media to publish the information. It might send the information to reporters or directly to consumers. It might try to persuade citizens to share the information with their friends. It might pay for billboards, advertisements, and newspaper articles. It will try to out-compete other accessible information to gain the attention of consumers.

In China, the central and local governments conduct information flooding by coordinating messages between in traditional and online media. Government propaganda documents describe a successful coordination effort as one where the local government contacted as many newspapers as possible to get them on board with supporting the construction of a controversial para-xylylene (PX) plant (National Academy for Propaganda Cadres 2011). The government found scientific studies that argued that the plant would have minimal environmental damage and made them available to these newspapers. They hired experts who supported their opinion to post in online forums. By encouraging all the newspapers to print the same perspective and posting on forums themselves,

they overpowered and pre-empted online opinion opposing the plant. In Chapter 5, I will discuss the effects of such coordination across newspapers on the spread of information in China.

Information Friction

When the government or organized group would rather not distribute the information, it can employ a strategy I call *information friction* to increase the cost of information for both the media and the public by making the information more disaggregated and inaccessible. The group creates small frictions that discourage the consumer or reporter from accessing or aggregating the information. It might literally increase costs by requiring a fee for data access or for interview access. It might draw out the time necessary to collect the data.

In China, central and local governments often create information friction by persuading or requiring websites to remove information about an event. Government propaganda documents describe the importance of closely monitoring online information about protest events so that the government can make quick decisions about what information to take off the web and how to respond to the event. They describe the importance of persuading websites, particularly local websites, to moderate forums (National Academy for Propaganda Cadres 2011). Doing so makes information about the event more difficult to access, particularly for consumers in the locality of the protest, as consumers will have to spend longer searching for it online. In Chapter 4, I will describe these efforts of information friction in more detail and estimate how these frictions influence the spread of information about protest events in China.

Fear

In situations where the government or organized group would prefer not to distribute information, many scholars have suggested that it can effectively prevent the spread of information through in-

timidation (Keohane and Nye Jr 1998; Boyle 1997; Hughes and Wacker 2003). I contend that in many cases, fear is not the most effective strategy of stopping the spread of information. First, much of the information that the government or organized group would rather people remain hidden is not already known to many people. Therefore, intimidating people by telling them not to spread that information might actually uncover the information the government or interest group would rather people not know, the opposite outcome intended by the group.

Second, fear sets the agenda by making the undesirable issue more salient. Since fear is a salient emotion, it rises to the top of a consumer's consciousness, and she will likely prioritize that information rather than repress it. Generating fear makes the fact that the group is intimidating the media or public well-known and highly available in the citizen's mind. Instead of using fear, I find that groups often try to use information flooding to distract consumers of information, forcing other, more desirable news stories to the front of their awareness.

2.2.5 Preferential Access to Information Undermines “Free” Information

One of the main determinants of whether organized interests and the state can implement information friction or flooding is the degree to which the information they collect is *preferential* (only available to that particular group), or *public* (generally available). Imagine two different cases. First, as I describe in Chapter 1, the Chinese government tries to make information about air pollution more costly by flooding the information environment with its own, falsified data, and creating friction in the form of the Great Firewall for U.S.-collected air pollution data. However, its efforts are thwarted because air pollution data is not preferential – citizens cannot help but notice air pollution and can collect data on air pollution. Although in this case consumers did not begin to collect the information immediately, friction and flooding, were not ultimately successful because other organized groups began collecting the same information themselves and distributing it.

In a second example, imagine that the Chinese government is trying to hide information about high-level corruption within the Chinese government. This information is very difficult for the public or any other group other than the government to collect; it is highly preferential. It would be much easier for the government to create information friction on this topic, as the information would be extremely costly for the public or the media to collect themselves, and the public would likely not even know what to look for.

Preferential access to information means that obtaining the information is higher cost for the media, and so the media are less likely to make it accessible. The degree to which information is preferential may be affected by the amount of control the organized group has over the information and whether special expertise is required to interpret the information. The slow outbreak of a disease, for example, is likely to require data collected over the long term by hospitals, and the analysis of epidemiologists. However, disasters like the Chengdu earthquake in 2008 are sudden, spectacular, and relatively easy for witnesses to understand. While some of the details of the government's response to the earthquake may be preferential, information about the central narrative and number of people who died may be inferred from eye-witness accounts.

2.2.6 Competition Among Groups Creates “Free” Information

The other axis that determines the success of information friction and flooding is the degree of competition that the organized interest faces. If an organized interest is the only group collecting information, it will have almost complete control over the cost of the information to the media and the public. If, however, a competing interest group also collects this information, the second group may make this information available at a much lower cost. Such competition can eliminate the power of the organized interest to control the spread of the information it collects.

Competition can affect information accessibility even when the information is preferential. In

the earlier example, even where the Chinese government tries to hide preferential information about high-level corruption, the information could be leaked by a rival faction within the government elite. For example, recently the U.S. press revealed the wealth of Wen Jiabao. Many suspect that the U.S. papers received the information for this story from Wen's rivals within the CCP (Sisci November 1, 2012). Political competition within the government can reduce its control of information, if one group has an incentive to reveal the information to the media.

Competition among groups also means that the independent media has more sources from which to draw when gathering information about a story. If only one group has access to preferential information, the group has a monopoly on the story, and may share only the aspects of the story it wants to reveal. If multiple groups have access to the information, however, the media and to some extent the public may be able to access some information from within one and some within another. This could result in a more complete story and may also allow the media corroborate details by comparing the two accounts. I detail how political competition influences the efficacy of censorship and propaganda in China in Chapters 4 and 5.

2.3 Structure of This Manuscript

In this manuscript, I test the above theory within China, demonstrating that the mechanisms behind information control described above can explain to the influence of Chinese censorship, intimidation, and propaganda on the spread of information. While China, with half a billion internet users, is important in its own right, I choose China because any successful theory of the spread of information under manipulation must hold true for the country with the most sophisticated selective censorship regime in the world. China is a good test case also because of the variety of information control strategies employed by the government.

However, the implication of my finding – that friction and flooding, not fear, are the main mech-

anisms behind online information control – is that many governments and other groups use similar methods of control and therefore the theory of information friction and flooding spreads much farther than China, authoritarian governments, and even governments themselves. In ongoing research, I apply the theory to other groups and the U.S. government, examining how information manipulation influences the spread of information in the U.S., and compare these findings to the Chinese case.

Chapter 3

Fear and the Paradox of Visible Censorship

The main debate within the current literature on information control is whether the internet benefits the state at the expense of society at large (Zheng 2007*b*). While some argue that the internet is difficult to control (Yang 2009; Ferdinand 2000), many maintain that fear is still the primary way through which the state can influence society in the age of information (Kalathil and Boas 2010; Lessig 1999). In this chapter, I address the previous literature by estimating how fear influences the consumption and production of information in China. I use novel experimental methods and a new longitudinal dataset of bloggers to examine individual-level behavior of both consumers and producers of social media. I find that the awareness of censorship has a very limited influence on the spread of information, and in many cases may backfire on those trying to create fear. I find much less self-censorship than previously speculated in the literature: netizens are *more* likely to read a blogpost if they know the topic has a history of censorship, and authors of blogposts have only muted topical reactions to censorship.

In China, censorship causes anxiety because of the history of government reprimands for the production and consumption of information. The Chinese government not only jails bloggers and other social media users on a regular basis, but also has a history of information policies that

have an explicit purpose of inducing fear in *consumers* as well as producers of information. For example, during the Maoist period, owning a banned book or having a picture of a purged leader could result in re-education in a labor camp (Jin 1999). Recent jailing of bloggers in China have generated significant amounts of news, and regular social media users have been arrested. While most of the arrests have focused on high-profile bloggers, a middle-schooler was detained after he claimed the police in Gansu province had been complicit in a death at a karaoke bar (Tang September 30, 2013).

The observation of censorship might signal to bloggers which topics the government deems objectionable. When a social media user observes censorship, she might decide not to read or write about a particular topic because she fears government reprimands. This is a direct mechanism through which fear of the state could function: because people observe censorship, they might decide not to inform themselves or others about a particular issue, thus slowing the spread of information.

Fear is a different mechanism than the *information friction* described in Chapters 2 and 4, which is an increase in the cost of access and does not require an awareness of censorship. Information frictions that decrease accessibility, such as the selective removal of blogposts online, website blocking, or denial of service attacks, may be invisible to the public, but raise the costs of access sufficiently so that consumers may never obtain that information. These netizens may be completely unaware that censorship is occurring, but their consumption of information will still be affected by it.

Many previous scholars have argued that awareness of sensitive topics is the primary mechanism through which censorship is effective in stalling the spread of information in China. Intimidation of the news media is probably the oldest form of censorship, and the government's monopoly on force makes it easy for those in power to inspire fear, regardless of technology (Keohane and

Nye Jr 1998). For example, Boyle (1997) reasons that fear of unknown and sporadic controls of the internet induce self-censorship. Hughes and Wacker (2003) argue that the perception of surveillance in Chinese society and a few high-profile arrests have more impact than actual censorship in stopping the spread of information.

In this chapter, I demonstrate that fear is an ineffective censorship tool within the Chinese blogosphere because it draws attention to the very topics that the government intends to keep secret. The awareness of censorship *alerts* bloggers to what the government would prefer to hide. This is the paradox of visible censorship: it raises the profile of prohibited topics and brings government/society conflicts to the foreground, rather than obscuring them.

Intimidation can especially backfire when censorship is opposed by netizens. Even though censorship can easily produce fear, it is clearly attributable to intentional government action. When a person sees that her post or another post has been censored, she knows that the government objects to its content, and that continuing to write on that topic could have consequences. However, if the person opposes censorship, the anxious reaction might be outweighed by dissatisfaction with government censorship policies, and might instead reinforce her negative views of the government, inspiring her to continue writing as an additional act of protest.

Censorship could also signal government weakness. Censorship is an indication that the government has something to hide, and therefore may signal to an observer that the government is weaker than previously believed. For example, Hassanpour (2011) presents evidence that the complete internet blackout in Egypt during the Arab Spring undermined government legitimacy because of its observability. Signs of weakness could embolden writers to continue writing, despite the dangers of doing so (Huang and Li 2013).

The internet has created avenues for making information difficult to access that do not require consumers of media to be aware of censorship. Highly organized groups and governments can con-

control the costs of information by making information relatively easy to access, or relatively difficult to access and through these means prioritize media for online users. This type of friction and flooding can manipulate consumption of media without being obviously government driven, therefore achieving both objectives of controlling the spread of information and maintaining government legitimacy.

Intimidation is more effective when it targets high-profile bloggers, but is invisible to average bloggers and consumers. If bloggers with large numbers of followers self-censor, it creates a secondary effect of friction; the fewer blogposts are written, the more difficult it is for consumers of blogs to find information about that topic. Thus, intimidation targeted at high-profile bloggers, while undermining the legitimacy of the government in the eyes of these bloggers, does not undermine the legitimacy of the government in the eyes of consumers of blogs who are not aware of censorship.

On the other hand, targeting consumers of blogs does not create further friction. Consumers of blogs have no secondary influence on the population; deciding not to consume information about a certain topic only influences the reader, not others. Therefore, because fear instilled in consumers cannot create friction, this strategy only reduces the amount of information that one person has, at the same time as it works to reduce the legitimacy of the government in the eyes of that one person.

In this chapter, I measure self-censorship in both consumers and producers of blogposts in China. I show that censorship piques the interest of consumers of blogposts, causing them to seek out more information, instead of less, on that topic. Bloggers exhibit small and not substantially significant levels of self-censorship, writing slightly less about political topics after censorship than before censorship. They do not slow their production of blogposts after censorship, and are no less likely to be censored directly after censorship.

3.1 Data

I use two datasets to understand how the awareness of censorship influences consumers and producers of blogposts and in turn the spread of information. In each test, my aim is to study *typical*, rather than *high-profile*, producers and consumers of blogposts in China. In the first test, I gather posting data and censorship data on a set of bloggers over a two-year period. I record the topical content and sentiment of bloggers before and after censorship. I study how censorship changes the topics bloggers chose to write about. In the second test, I hold information friction constant and randomize the anxiety about censorship in an experiment with consumers of blogposts in China. I study how the observation of censorship influences participants' subsequent consumption of blogposts.

3.2 Limitations

While these tests both seek to understand different aspects of how fear influences the spread of information, they have some limitations. First, I study the effects of censorship on individual, typical bloggers. I cannot observe how people who have never blogged are influenced by censorship, and I do not study how very famous bloggers are influenced by censorship. It could be that some people do not blog at all because of censorship, or that high-profile bloggers are more affected by censorship than typical bloggers. Second, since I cannot follow the online actions of typical consumers of blogposts, I study fear in an online experiment with subjects in a lab setting. While I try to make the experimental setting as natural as possible, not all of the results may be externally valid. Last, my studies are limited to the Chinese context, and, although China has one of the largest programs to selectively censor, these findings will not apply to every other country. Each of these limitations could be addressed in future studies of self-censorship.

3.3 Test One: Do Bloggers Change Topics After Censorship?

My first test of how fear influences the efficacy of censorship studies the writing patterns of bloggers. Unlike consumers of blogposts who observe censorship only when where links are broken or forums have missing threads, producers of blogposts are more likely to become aware of censorship. Particularly because censorship typically occurs within one day (King, Pan and Roberts 2013a), and therefore bloggers' most recent posts are likely to be the posts that have been removed, bloggers may notice that their blogpost has gone missing.¹

Bloggers in China have historically been arrested or harassed by the government as a result of the content that they create and therefore are likely to suffer graver consequences for their actions than consumers of blogposts. Famous activist and artist Ai Weiwei was arrested in 2011 ostensibly for tax evasion, but most observers believe the true reason for his arrest was posting a comment encouraging the Jasmine revolution in China. More recently, in the beginning of 2013 a handful of bloggers were arrested in a crackdown on blogging in China. The national government has also implemented new rules prohibiting "rumors" and certain types of political discussion that give the government wider latitude to arrest and prosecute bloggers. The new law states that if a rumor or other illegal post is re-shared or viewed enough times, the blogger could be jailed for up the three years. This law puts any blogger at risk of being jailed, including the middle-schooler discussed above.

Overall, one might expect that bloggers would be deterred from writing about sensitive political issues after experiencing censorship. The awareness of censorship signals the type of information the government finds objectionable, and therefore might encourage bloggers to shift focus away from that topic to a less objectionable one. They might also change the way they portray the

¹In increasing numbers of cases, internet content providers try to conceal censorship from bloggers, by allowing them to view the censored posts if they are logged into the blog. This practice, however, is still relatively rare, and was not the case in the website used for this study.

government, if they are afraid of being arrested.

3.3.1 Data and Methods for Testing Blogger Self-Censorship

To study how the awareness of censorship influences bloggers, I sampled 2,000 bloggers whose blogs were hosted on a particular internet content provider in China that appeared in the King, Pan and Roberts (2013a) dataset. I chose this sample because I had collected data on all of the blogposts these bloggers had written since 2010. The blogposts were collected before the Chinese government was able to access and delete them, so they reflect all posts each blogger wrote. These bloggers also represented a wide range of characteristics, some had very few followers, others had more followers, and of all ages and genders. I then returned to each of the 150,000 blogposts to see which had been censored or removed by the internet content provider. Therefore, for each of these bloggers, I had the content of what they had written and when each of the bloggers had been “treated” with censorship.

To study how censorship changes the behavior of bloggers, I focus on estimating the within blogger change in topics before versus after censorship. In the period immediately before the blogger experienced censorship, did the blogger write about different topics than during the period immediately after the blogger experienced censorship? Because I am interested in changes in topics within bloggers, I removed all bloggers who had never experienced censorship. Topical variance among bloggers is quite high, and therefore a comparison between bloggers who had been censored and those who had never been censored would reflect which topics were censored, not how bloggers are influenced by censorship. Although the types of topics that are censored is an interesting question in itself, I reference the reader to King, Pan and Roberts (2013a,b) for discussion of this question and instead here focus on how individuals are affected by censorship.

I also removed all bloggers whose blogs had been removed altogether, as I could not tell if they

were missing due to government intervention or the blogger's own choice. After sub-setting, 516 bloggers of interest remained. From each of these 516 bloggers, I randomly selected a censored post. I then selected the five blogposts directly preceding the censored posts, and the five blogposts directly following the censored post.² Altogether, my dataset included with 5,199 blogposts from 516 bloggers.

3.3.2 The Structural Topic Model

To analyze the blogposts, I used a new method for text analysis called the Structural Topic Model (Roberts, Stewart and Airolidi 2013; Roberts et al. Forthcoming). The STM is based on Latent Dirichlet Allocation (Blei, Ng and Jordan 2003), which extracts topics in an unsupervised way from unstructured texts. However, STM allows for the inclusion of arbitrary covariates into the model for topic estimation, allowing better estimation of topics for documents that are not independent. Since I am interested in how a particular covariate (which blogposts were written after censorship) is related to topic choice, this model is ideal for the research question.

STM is also an appropriate choice for this application because it permits the inclusion of blogger-specific fixed effects and thus allows within-blogger estimation of how bloggers change topics in reaction to censorship. I include blogger-specific fixed effects and an indicator for whether the blogpost was written before or after censorship in the STM model.³

The Structural Topic Model estimates a variety of quantities of interest from the text data that are relevant to this analysis. First, it estimates a fixed number of topics, represented as a set probability distributions over the vocabulary within the corpus. A topic can be represented as the most likely words from each of these distributions, words that cluster together within documents

²If the censored post had fewer than five posts preceding or following it, I took the maximum number of posts that were available.

³I also scraped a variety of metadata about the blogger, including number of followers, age, and gender, and included these metadata within robustness checks, but the metadata had no significant outcome on the overall results.

and form coherent concepts. STM also estimates correlations between topics, indicating whether some topics are more likely to be used together in the same document. Last, STM estimates topic proportions, or the estimated proportion of each document that is about a particular topic. Using the topic proportions, I can compare the distribution of topics in blogs written before and after censorship. If, after censorship, bloggers shift away from sensitive topics, become less critical of the government, and delay posting, we would conclude that they are self-censoring. If bloggers exhibit no topical or sentiment shifts after censorship, we would conclude that the awareness of censorship does not deter bloggers from writing, and that any anxiety produced by censorship is perhaps offset by countervailing considerations or emotions.

3.3.3 Results

I estimate a 50 topic model on 5,199 blogposts. Topics could be broadly classified as either relating to politics, or relating to personal life. An overview of the topics and their prevalence within the dataset can be found in Figure 3.1. Political topics include a range from party politics to international relations. Personal topics range from online diaries detailing what the blogger did that day, to love, family, and work. Other types of topics include computers (how to download a file, what operating system to use), art (celebrities, poetry, etc), political history (from Ancient Chinese history to Mao), or economics (exchange rates, small business, and the stock market). An individual blogger is likely to talk about several different topics throughout the time period in which they are blogging.

Figure 3.2 shows a correlation plot of the topics estimated. A line between topics indicates that those two topics were positively correlated. The correlation plot shows the main axes of discussion among bloggers. The topics are sub-classified into larger topical themes, including government, economics, history, art, and personal topics. Personal topics and art are on the left side of the plot,

while the more political topics including government, economics and history are on the right side of the plot.

The question I aim to answer is: are bloggers more likely to switch to more personal and less political topics after censorship? In order to provide the most rigorous test of this hypothesis, I subset the data to bloggers whose censored post was estimated to be more than 70% about a political topic. This eliminates blogs where the censored post could have been removed for reasons other than government censorship, such as spam or topics that may have been removed by the blogger herself, providing a more rigorous test of the theory.

Overall, bloggers whose censored post is about a political topic are much more likely to discuss economic and political topics in all their posts than those whose censored post was not mainly about an economic or political topic. Figure 3.3 shows the most common topics for the subset of bloggers whose censored post was about a political or economic topic. I also looked in detail at the censored posts. The most common censored blogposts in this sample were posts related to the downfall of Bo Xilai and anti-Japanese protests. This is consistent with King, Pan and Roberts (2013a), who find that the Chinese government tends to censor blogposts about collective action.

I use this subset of bloggers to determine whether bloggers have different topic proportions before and after censorship. Bloggers do not exhibit a significant shift in topic for any individual topic. Figures 3.4 and 3.5 show the shift in topic on political topics, which we would expect to be most negative, and life topics, which we would expect to be most positive. All confidence intervals cross zero, indicating no significant shift.

Only when I aggregate political and life topics do I find that bloggers exhibit a statistically significant but substantively very small shift away from political topics after censorship, indicating a very small degree of topical self-censorship. There exists a significant shift away from topics related to politics. However, this shift is quite small: bloggers decrease the political proportions

Top Topics, Overall

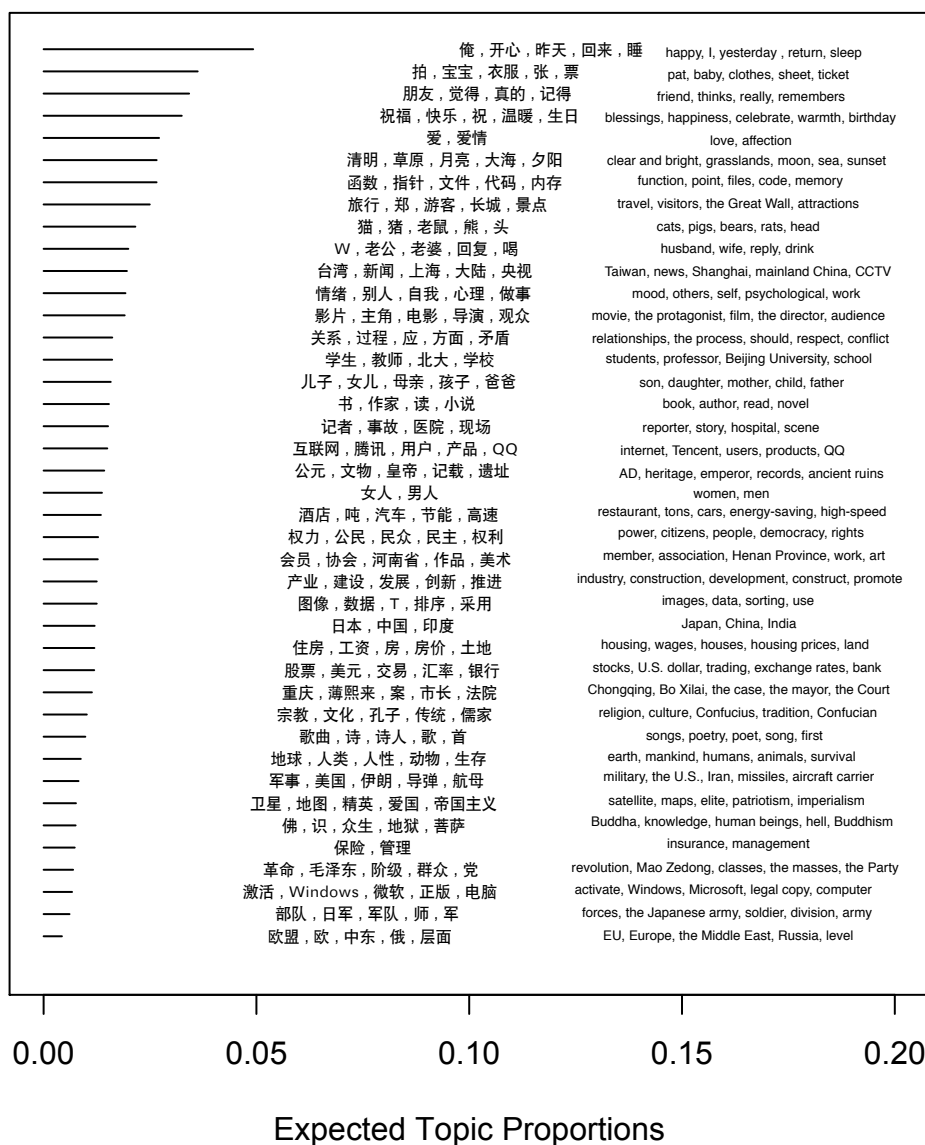


Figure 3.1: *Top topics across all of the blogposts within the study. The words presented are the most frequent words within each topic*

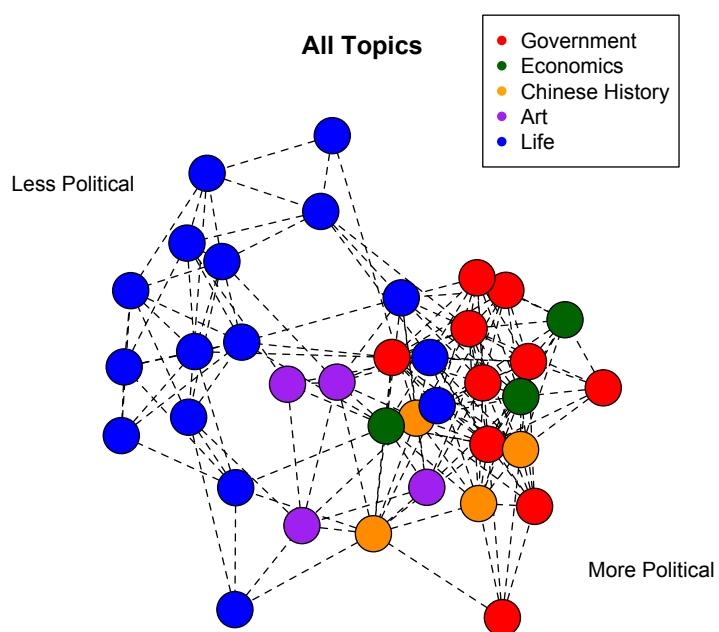


Figure 3.2: Topical clusters. A line between two clusters indicates that the topics are correlated. Colors indicate the type of cluster.

Top Topics, Political Subset

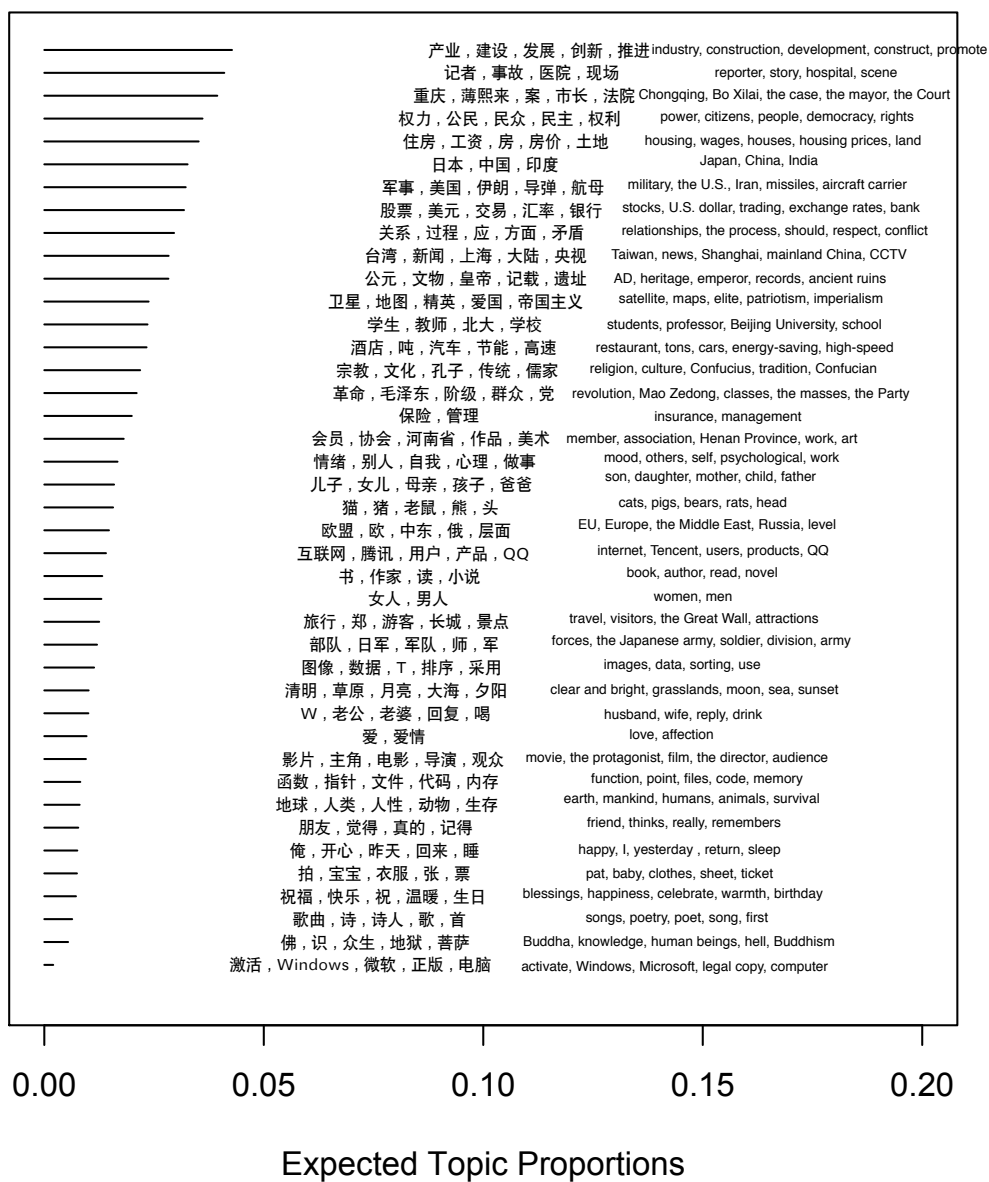


Figure 3.3: Topic incidence for bloggers whose censored post was political.

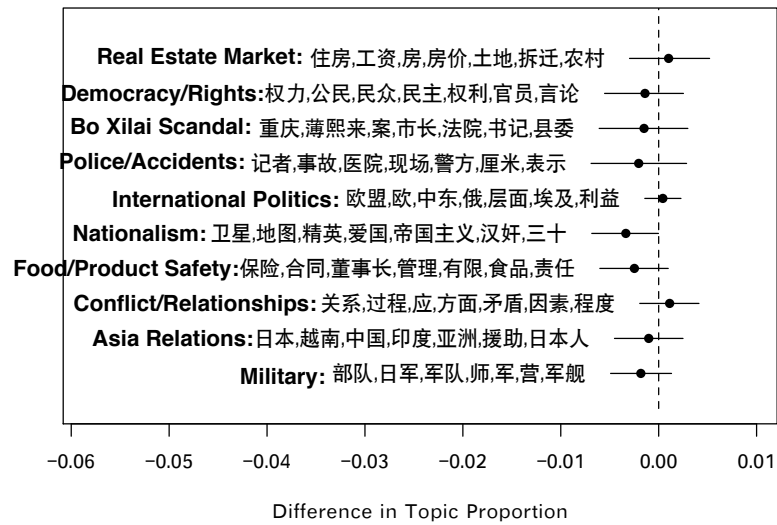


Figure 3.4: Shift in political topics, before and after censorship.

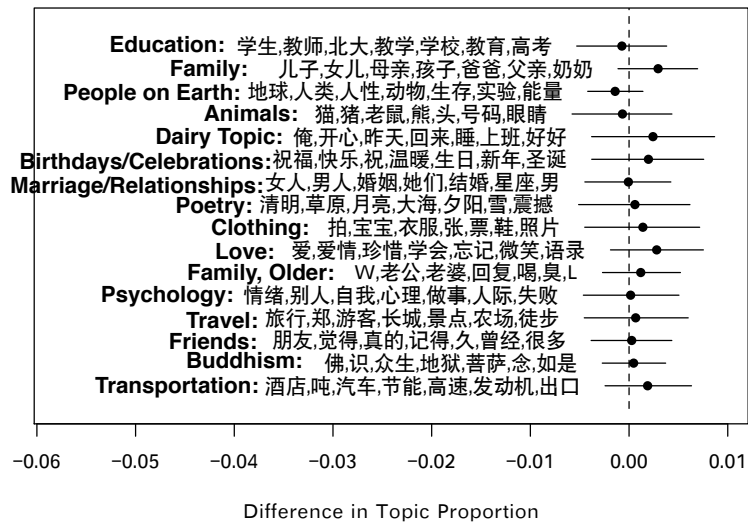


Figure 3.5: Shift in life topics, before and after censorship.

of their post 1% on average and increase post that discuss personal life by .1%. While fear of government backlash might influence the topic selection one in 100 bloggers, it does not seem to have a substantively meaningful influence and the effect may well be sub-conscious.

Censorship also does not seem to influence bloggers' other observable actions. If bloggers really were deterred from writing after censorship, bloggers might write at a slower rate after being censored, but a statistical test of the difference between the rate at which bloggers write before and after censorship is small and insignificant. The time between blogposts after censorship is on average one day longer than before censorship, but this difference is not significant, with a confidence interval of (-1.38, 3.78). Censorship also does not decrease the likelihood of future censorship, which one might expect if bloggers began to write about topics that were less objectionable to the government. On average, blogposts after censorship were 3% less likely to be censored than blogposts before censorship, but this difference is not statistically significant, with a confidence interval of (-.02,.08).

Censorship Does Not Change Blogger Opinions

While the STM can measure bloggers' topical shifts before and after censorship, it is not engineered to detect shifts in sentiment, which is often finer-grained than topics. To understand how sentiment changes before and after censorship, I hired two coders to read a random sample of blogposts. The coders indicated whether each post was critical, supportive, or neutral about the government. I then analyzed whether bloggers were more critical after censorship than before censorship. As shown in Figure 3.6, I found no significant difference between critical, neutral, and supportive categories before versus after censorship.

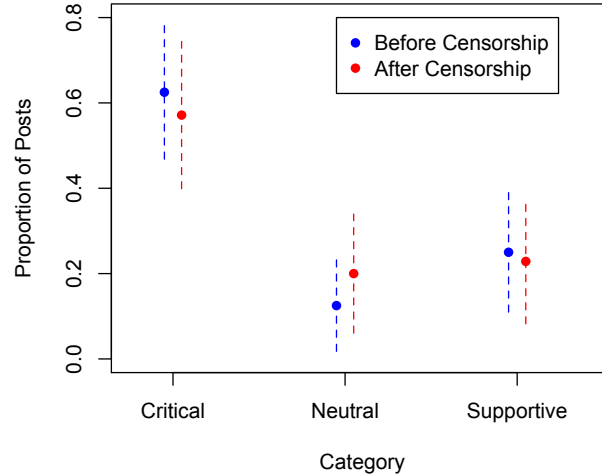


Figure 3.6: *Relationship between sentiment and posts before and after censorship.*

Robustness Checks

I conducted a variety of robustness checks to verify that these results persist across specifications. First, returning to each of the blogs, I scraped as much information as I could about the bloggers' age, gender, number of followers, and city. While much of this data is missing from bloggers' websites, I tested whether the topical reactions were related to the characteristics of the bloggers. Including these covariates did not change my results.

Second, it could be that the number of previous times a blogger has been censored affects whether they react to censorship. I ran a topic model using only the first blog censored for each of the bloggers, and another using only the last blog censored for each of the bloggers. Both topic models showed the same results: bloggers did not change the topics they discussed before and after censorship.

3.3.4 Implications for the Spread of Information

My finding that bloggers do not change topics or alter the sentiment of posts about the government in reaction to the awareness of censorship implies that fear produced by censorship does not slow the spread of information in China. Even when bloggers are aware of censorship, they are likely to continue writing about the same topics with the same sentiment as if they were unaware of censorship. Across half a billion Internet users in China, awareness of censorship is unlikely to stall the spread of information about a particular event, since consumers and producers of blogposts seem undeterred writing and consuming such information.

3.4 Test Two: An Experimental Study of Consumers of Blogposts

While censorship in China is less apparent to consumers of social media than it is to producers, occasionally a consumer of information will come across a censored page. One way observation of censorship occurs is through a link, when a page that the person is visiting links to the URL of a censored page. If the consumer clicks the link, she will be taken to an error page instead of the page with the original content. An example error page is displayed in Figure 3.7. Another way consumers of social media can observe censorship is within a forum or on Sina Weibo, the Chinese version of Twitter. When posts are censored within the forum or weibo, the content of the individual comment is removed, but the rest of the conversation still exists. The user will see a removed post where the original post once stood.

When a consumer comes across a censored page, she can often determine the topic of the censored post using context. For example, if the user clicked on a link to a censored page, the title of the link might indicate the content of the original blogpost. In a forum, consumers of



Figure 3.7: Error page indicating that a post has been removed.

information may be able to guess at a missing post's content from the surrounding discussion.

Consumers of blogposts in China often suspect that their online activities are subject to surveillance. Many internet content providers now require government IDs or a phone number for users to sign up for an online username. The government tracks online behavior of netizens who frequent internet cafes (Stevenson 2007). While a proposal to install tracking software nicknamed the “Green Dam” on all computers was abandoned after outcry from internet users, many local governments have nevertheless mandated similar software under different names (MacKinnon 2011).

In an age where online surveillance is common, if consumers of information come across censored blogposts or other types of social media, how do they react? If the consumer is significantly affected by fear of the state, particularly in light of the government's online surveillance, the experience with a censored blogpost could cause the reader to avoid further interaction with the topic. On the other hand, if censorship creates countervailing emotions, such as indignation or curiosity, the consumer may have an increased interest in reading more about the topic. These basic questions motivate the second test of how fear influences the efficacy of censorship: an experiment of the awareness of censorship's influence on consumers of blogposts in China.

3.4.1 Design

I conducted an experiment aimed at understanding how awareness of censorship affects consumers of blogposts with 150 students at Renmin and Qinghua Universities over the summer of 2013. Students were given a computer, either a laptop in a coffee shop or a computer in a lab setting and access to the internet. They were provided with a list of blogposts in a blog-aggregator. They were instructed to read blogposts that interested them, and told that they would be asked a few questions about the blogs they read after five minutes of reading. The subjects were aware that their actions online were being watched, as the description of the process indicated that their behavior online was being recorded, simulating an environment of surveillance.

The blogposts covered four different topics current events in China: protests surrounding the construction of para-exylene (PX) plants in Yunnan, protests in Hong Kong against the government, a scandal alleging that a Chinese Communist Party official's son, Li Tianyi, was involved in the gang rape of a woman, and blogs speculating that the Chinese economy would soon descend into an economic crisis similar to the one that occurred in the United States. Because this study was conducted in China, more sensitive topics could not be used, but pre-testing of the experiment suggested that these were all topics Chinese citizens would subject to censorship. All blogs within the study existed online and had not been censored at the time of the study.

Students were asked to read blogs that interested them and then told that they would be asked a few questions. They could judge which blogposts might interest them because the title of the blogpost was displayed on the main page of the aggregator, and each title contained information that would allow them to determine the post's topic. It was clear that they could not cover all topics or read all blogposts during the time allotted; therefore, they had to choose which posts and topics to read.

For the treated group, the first blogpost the subject clicked on from a randomly selected topic

would not link to the blog itself, but instead to the error page associated with the blog's internet content provider, indicating censorship. Instead of reading the blogpost, the student would then return to the blog aggregator and choose another blogpost to read. In the control group, all blogposts from that topic would link to the content of the post, instead of the error page.

The question of interest was: how does censorship of a blogpost influence the topic the reader selects next? Is the reader less likely to pick a blogpost on that topic because she has received a signal from the government that that topic is off limits? To measure this, I installed technology within the website to track the behavior of each individual. I could therefore observe when an individual clicked on a censored link, and if they did, which link they decided to click on after encountering the censored page.

A comparison between the group that encountered censorship and those that did not is the causal effect of the awareness of censorship on the reader's consumption decisions. If the censored page created mainly fear and anxiety, we would expect participants to avoid that topic in the future. If the censored page inspired curiosity or indignation, we would expect participants to be equally or even more likely to click on the censored topic.

3.4.2 Results

The treatment had its intended influence on the subjects. When asked what the error page indicated, most (two-thirds) of respondents said that the page was unavailable due to government censorship. In addition, those treated with censored posts updated their expectations of which topics the government censors. Those who were treated with censored posts about Li Tianyi, the son of a Chinese Communist Party leader, were significantly more likely to say that the government censored rumors about government officials than those not treated with censored posts about Li Tianyi. Similarly, respondents who were treated with censored posts about the Hong Kong or

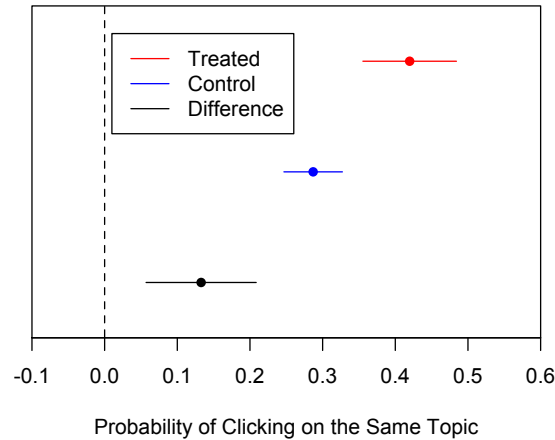


Figure 3.8: *Treatment effect of censorship on the probability on next clicking on a post of the same topic.*

PX protests were significantly more likely to say that the government censored information about protests than those not treated with censored posts related to protests.⁴

Censorship did not dissuade people from reading more about the same topic; instead, it made people more interested in the topic. Readers who came across censored posts were more likely to next click on a post on the same topic than those who came across uncensored posts, keeping the topic of the blogpost constant. Further, readers who come across censored posts were clicked on more posts about that topic over the course of the experiment than readers who were not treated with censorship (Figure 3.9). The results are strongest for respondents who thought that the error page was due to censorship, and insignificant for those who thought that the error was due to an internet error or to the removal by the blogger.

More tellingly, treated subjects had different views than control subjects about how much the

⁴Respondents were all debriefed after the survey was conducted and told that none of the posts within the experiment were actually censored. This project had full approval of Harvard's Institutional Review Board.

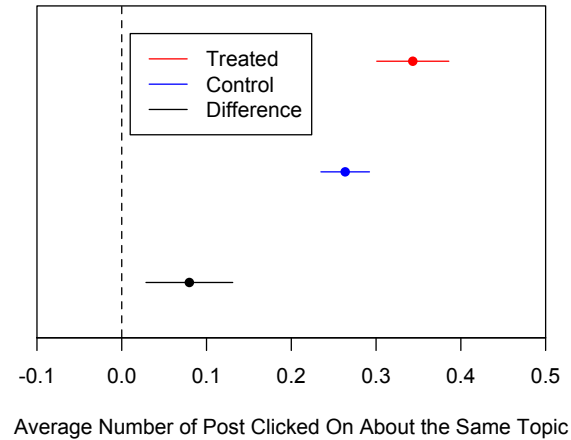


Figure 3.9: *Treatment effect of censorship on the number of posts selected on the censored topic after treatment.*

government should censor the internet. When asked about political censorship, including censorship of discussion forums, censorship of rumors, and censorship of “malicious information” (terms the government uses to describe the material it censors), treated subjects were more likely to say that they thought censorship should be reduced than control subjects. This difference in opinion suggests that the observation of censorship creates countervailing emotions that undermine the legitimacy of the government’s information laws.

The results imply that awareness of censorship on the part of consumers of blogposts does not dissuade these consumers from reading more about a topic. Instead, observing censorship interests readers in learning more about the topic, and increases frustration with government censorship laws. Censorship, in this context, may backfire and actually undermine government policy.

If readers are intrigued, not deterred, by visible censorship, observable censorship will increase the spread of information rather than slow it. Consumers are more likely to read about the topic, and

might therefore be more likely to share it with their friends. They might be angered by censorship and emboldened to oppose the government's censorship policy. The mechanism would therefore have the opposite effect than the previous literature would have expected.

3.5 Conclusion

If fear is not the way through which governments can stall the spread of information in online media, then through what mechanism, if any, can governments wield information control? While visible censorship, like the name would suggest, backfires on the motivations of its initiator, invisible censorship, unnoticed by consumers, can have significant effects. In the next two chapters, I turn to less visible censorship: information friction and information flooding.

Chapter 4

Information Friction and Protests in China

Soon after the introduction of the internet, scholars posited that the masses would be the beneficiaries of this new technology, at the expense of the powerful (Zheng 2007*a*; Browning 2002; Grossman 1995). The internet created transparency, providing minute-to-minute news on actions of governments, politicians, companies, and interest groups. The internet allowed citizens to communicate with each other instantly, creating a new forum for civil society that could spread signals of discontent and organize action against the government, before bumbling censors could access this information (Yang 2009). And the internet thwarted government and interest-group led efforts of censorship because with the internet, nothing could be kept secret (Taubman 1998).

However, the perception that the internet has made all information accessible is fundamentally flawed. While the internet has made a lot of information easy to find, much information is still disaggregated or difficult for the public to access. Market research provides evidence that internet users are for the most part lazy consumers of information, and will only consume the most accessible information (Hoelzle January 2012). Governments and interest groups, who are the main actors gathering and providing access to information, will determine what information consumers are most likely to read.

In this chapter, I focus on *information friction*, or the act of making information more disaggregated and more inaccessible to the public. Governments and interest groups are the most likely groups to employ an information friction strategy. First, these groups are willing to use resources to make information difficult to find. They will build infrastructure or pay individuals to make information less accessible. Second, these groups are often the primary sources of information. Groups and governments collect data on themselves, and therefore often are the gatekeepers to such information and can easily make this information more or less costly to access.

In China, the government employs an extensive censorship infrastructure to make undesirable information more costly to access. In particular, in this chapter I focus on content filtering, which is when government censors remove individual blogposts from the internet. The Chinese government uses content filtering to remove information about protest events within China (King, Pan and Roberts 2013a). By removing individual accounts of protest, the government is creating information friction – information about the protest events less accessible, as the information is more difficult to find online.

In this chapter, I collect hundreds of thousands of blogposts about three different examples of protest events in China before the Chinese government reads and censors them. I measure censorship of each of these postings and take data when available about the geographical location associated with the social media posts. Leveraging variation in censorship across websites, location, and time, I demonstrate that information friction, despite its imperfections, reduces the number of new blog postings about an event and reduces the geographical spread of information about that event.

In doing so, I also unravel the conditions under which information friction is more and less effective. I show that when government policy toward a protest event is more monolithic, content filtering is high and consistent across websites, slowing the spread of information. When, however,

various groups within the government have different opinions on the extent to which a protest event should be censored, censorship is lower and less consistent across websites, and information about the protest event spreads more quickly. This supports the theory that competition between groups or within governments can thwart information friction efforts, as some groups will work to make information less accessible, but others will cancel their efforts out by packaging or promoting it.

The chapter proceeds as follows. First, I discuss the connection between theory of information friction and content filtering, focusing on how the Chinese government uses content filtering to stop the spread of information during protest events. Next, I describe three different types of protest events in China: protests against the Chinese governance of Tibet through self-immolations, environmental protests aimed at industrial projects supported by local governments, and nationalistic protests against Japan. Within each protest event, I use variation in censorship across websites, time, and geography to provide a detailed quantitative analysis of the influence of content filtering on the spread of information. I find that websites, periods of time, and geographies with higher levels of content filtering witness a reduction in the spread of information.

4.1 Content Filtering Creates Information Friction in China

The Chinese government is well situated to create information friction. First, the government has invested enormous resources in the infrastructure necessary for censorship. The Chinese government's censorship program is one of the most extensive in the world, employing thousands of people and operating over hundreds of different websites (King, Pan and Roberts 2013a). Second, authoritarian governments notoriously have difficulty gathering information about the public (Lorentzen, Landry and Yasuda 2010) and therefore the Chinese government monitors public opinion online and therefore is already following social media (Denyer August 2, 2013). Its extensive and costly censorship infrastructure, in conjunction with the fact that it is already monitoring so-

cial media, means it can react very quickly and early to circumstances when it hopes to create information friction by removing content about a particular event.

The main result of this extensive censorship and monitoring apparatus is not keeping information completely secret, but making objectionable information less accessible in order to create significant frictions for netizens who try to access it. Famously, the Great Firewall, which blocks foreign websites in China, makes accessing foreign information more costly for the public because it requires a proxy or a Virtual Private Network (VPN) to circumvent. A less known example, search filtering, reorders search results such that information the government would rather not people read about receives a lower search rank. While information about these topics still exists online, search filtering frustrates the user by making search engines less useful in locating incriminating information. The government also slows down websites, like Google mentioned in Chapter 1, that are likely to contain information they would rather the public not access. Indeed, the result of each of these tools is not the complete destruction of information, but rather the creation of slight barriers to access that may frustrate consumers – information friction.

This chapter will focus primarily on only one of the Chinese government's censorship methods: content filtering, which is the selective removal of blogposts online in China. While the Great Firewall blocks foreign websites, the vast majority of social media posts in China are written on Chinese-owned internet content providers (ICPs). These website provide blogging, micro-blogging, forum, or bulletin board services, allowing Chinese citizens to express and share their views online. There are more than 600 million internet users in China, and the typical user has accounts with several different services (DeSilver December 2, 2013).

The Chinese government devolves responsibility of content filtering of social media posts to each of these individual internet content providers. When a user posts something that is objectionable to the government online, the website is responsible for removing this material from the

internet. If government censors find too much objectionable material on a website, they have the authority to shut the entire website down. Under threat of extinction, internet content providers employ thousands of censors who remove content based on weekly or daily directives from government agencies.

Large-scale studies of content filtering in the past have shown that the government focuses the efforts of censors on removing posts related to collective action events. King, Pan and Roberts (2013*a,b*) show that surprisingly the government does not target criticism, but rather removes all posts related to collective action events regardless of their support or criticism of the government. Bamman, O'Connor and Smith (2012) show that censorship focuses on social media posts that are geolocated in more restive areas, like Tibet. The aim of government censorship seems to be to stop information flow from protest areas to other parts of China, or to prevent people from knowing about protest events. Since large-scale protest is known to be one of the main threats to the Chinese regime (Chen 2012), success for the Chinese censorship program is preventing the spread of information about protests in order to reduce their scale.

Despite extensive content filtering, if users were motivated and willing to invest time in finding information about protests, they could overcome information friction to find such information. First, information is published online before it is removed by internet companies. There usually exists a lag of several hours if not a day or two before content is removed from the internet. Therefore, even the most objectionable material will spend a period of time online and will be available to the public before it is removed.

Second, internet content providers will occasionally miss posts related to an event. Automated methods of content filtering are not sophisticated or very successful, and therefore much of content filtering is done by hand. Since censors cannot read every post on the Chinese internet, they may miss a fraction of objectionable material. Netizens who want to discuss a particular event may also

find ways to trick the censors, either by finding websites that are less carefully watched (talking about politics on a dating website, for example), or by finding phrasing that makes posts about the topic difficult for the censors to find.

Last, if the event is reported in the foreign press, internet users could access the information by jumping the Great Firewall, using a VPN. Even if Chinese internet content providers remove information about these events, Twitter, Facebook, or other blocked social media companies will often contain information about these events. These companies do not filter content at the instruction of the Chinese government, and therefore, as long as a user has the knowhow and money to access a VPN, they can access the entire unfiltered foreign web.

However, despite the possibility of accessing information in the face of content filtering, the slightly increased costs of information due to content filtering reduce the probability that netizens who have a preference for easily accessible information will come across information about protests, and therefore have a significant influence on number of people who know about an event. This is primarily because content filtering reduces the prevalence of information on the Chinese internet, meaning that fewer people run into this information while searching the web, it is less likely that the information will go viral. Content filtering particularly focuses on bloggers who have large followings (Zhu et al. 2013), thus reducing the spread of information about the protest.

Content filtering not only decreases the accessibility of information, it also can disaggregate accounts of an event, making the full picture difficult to piece together for netizens. Users cannot tell whether the lack of information online about a particular event is due to government censorship or to general disinterest in the event. If “no one” seems to be talking about protests in Inner Mongolia, for example, even people who know about the protest may assume the protests do not have widespread support. They may assume that the protest was started by radicals, and that people they follow online, their “friends”, do not support these people. Unless they subsequently

spend time searching and aggregating multiple accounts of the event, netizens may not realize the importance of an event even if they simply happen upon one post related to it.

Last, even if users can access unfiltered information on foreign websites, very few Chinese netizens choose to use a VPN to jump the Firewall. In a survey of the participants in the experiment described in Chapter 3, only 8% of the participants said they frequently jumped the Firewall.¹ When I asked the other 92% of participants why they didn't jump the firewall, only five said that it was because of government regulation; the majority said that it was because jumping the firewall was "too annoying". Many also said that there was no reason to jump the Great Firewall – what useful information was there behind it? The combination of the small costs of jumping the Firewall and the uncertainty of its benefits is sufficient reason to deter large numbers of netizens from using VPNs.

All of these small costs add up to create a multiplying influence on the inaccessibility and disaggregation of information about the protest. The more content filtering, the fewer people happen upon the information and the fewer new postings occur online. The fewer the new postings, the fewer people know about the event, and those who do are more likely to think the event has fewer followers than they would if there was more discussion online. The fewer people who know about the event, the fewer people know there is any reason to jump the Great Firewall and are therefore unwilling to pay the small costs of a VPN. The smaller the number of social media posts that are related to the event, the fewer people out protesting, and the smaller effect the protest has on governance.

¹Since students are the most tech-savvy part of the population, we would expect a much lower percentage to use a VPN in the broader Chinese population.

4.1.1 Internal Government Conflict Reduces the Efficacy of Censorship

If content filtering works to create information friction, as I suggest here, then the more content filtering about protest events, the more effective censorship will be in reducing the spread of information and the scale of protests. We would expect a negative monotonic relationship between content filtering and the spread of information. However, internet content providers sometimes receive conflicting content filtering directives from a variety of different government institutions. In these cases, content filtering should be less consistent and less effective than when messages from government institutions are very consistent.

Internet content providers receive directives from various government agencies for what content they are required to filter. All ICP's receive directives from the central government about censorship, but some also receive local government directives associated with the location of the company. In addition, if companies have government connections in a particular part of the bureaucracy, they will receive and censor material from that institution.

As a result, coordination of censorship across websites in China is sometimes difficult because of conflict within the government. Shih (2008a) shows that factional conflict frequently shows up in newspapers, where particular provinces more aligned with the leading faction will more frequently reprint the leading faction's political slogan. In addition, competition between local governments for promotions may create incentives to shirk on censorship. A protest in another province would reflect poorly on the leader of that province, so a competing provincial leader could have incentives to delay censorship. Local government officials have also been known to pay censors to remove material they find objectionable when the central government will not issue a directive to do so (Zhu April 21, 2014). Such within-government competition creates opportunities to measure how competition between groups influences the efficacy of both information friction and information flooding.

The more consistent government censorship is over time and space, the more effective it is in stopping the spread of information. Competition among groups or even simply a lack of coordination across websites can make information friction less effective. However, even if censorship directives are not consistent across the whole country, it does not mean content filtering is completely useless. As I discuss more below, even when there is widespread discussion of an event, content filtering focused on a particular website or geographical location can reduce discussion on that website or in that geographical area.

4.2 Three Examples of Protests in China With Varying Levels of Censorship

In this section, I describe the three different examples of protest events I study within this chapter: self-immolation protests against Chinese governance in Tibet, environmental protests against industrial projects with local government backing, and anti-Japanese protests in the fall of 2012. While the government targeted each of these protests with content filtering, the consistency of censorship across websites varied across events. Censorship over self-immolation protests was monolithic and consistent, censorship over environmental protests was originally targeted in the locality, and only later became consistent over the whole country, and censorship over anti-Japanese protest exhibited variation across websites. Such variation allows me to estimate the effect of content filtering on the spread of information, and piece out how competition between groups influences the efficacy of information friction.

4.2.1 Monolithic Censorship: Self-Immolations in Tibet

Between March of 2011 and July of 2013, 120 Tibetans self-immolated within China, the majority of whom died after setting themselves on fire. While the exact reasons for the self-immolations are unknown and probably vary substantially by person, writings by self-immolators call for Tibetan independence, greater Tibetan autonomy, or the return of the Dalai Lama, all policy stances that the Chinese government opposes (Regalbuto 2012; Han 2013).

These protest events represent a huge problem for the Chinese government, as this recent spate of self-immolations follows large-scale protests in 2008, where thousands of young people protested in Tibet, some waving the Tibetan flag and calling for independence (Greve 2013). The self-immolations themselves are also frequently followed by larger-scale protests. Since the maintenance of harmony between ethnic groups and fighting independence movements is central to Chinese national security policy, the immolations are a direct challenge to the Chinese government. Self-immolations in other countries also have caused political upheaval; for example, Mohamed Bouazizi, a Tunisian businessman, is credited with sparking the Arab Spring in 2010 after self-immolating in political protest (Lotan et al. 2011). A self-immolation protest by a young factory worker Jeon Tae-il in November of 1970 began a larger movement for workers rights in South Korea, that eventually culminated in democratization (Armstrong 2006).

To discourage protest events, the government quickly responds to self-immolation events, removing the self-immolator as quickly as possible and increasing police presence to prevent protest events in the aftermath. Police often punish villagers and families of Tibetans who have self-immolated to discourage future events. Monasteries are often surrounded with police force since many of the self-immolators are Buddhist monks (Greve 2013).

Accounts and news of Tibetan self-immolations are uniformly and quickly censored on social media websites. No locality has vested interest in Tibetan independence, and therefore, all agree

with government policy and follow it. This is aided by the repetition of such events within China: local governments will automatically censor accounts of these events, since all similar events in the past have been censored.

Information about self-immolation events is also more excludable than other types of protest events. Such events often occur in rural or semi-rural areas within Tibet, where much less infrastructure and internet access exists than in other parts of China. As a result, government authorities and police are likely to be some of the first to encounter such events and therefore have more control over the dissemination of information about those events. This provides a contrast to other protest event in China that occur in urban areas, with thousands of people who often have access to the internet through their mobile phones, and therefore third party accounts of the event are more numerous.

4.2.2 Local Conflict: Environmental Protests Against Local Projects

Over the past five years in China, NIMBY, or “Not in My Backyard” protest movement have become increasingly common. These locally-organized protests opposing the construction of industrial plants because of environmental concerns have been successful in generating government concessions, and in some cases have caused local governments to suspend plans to build the plants. The two protest events I focus on occurred in July of 2012: one in Shifang, a city within Sichuan province and the other in Qidong, a city on the coast of Jiangsu province.

Shifang Copper Smelting Plant

The Shifang protest occurred between July 1 and July 3, 2012, when protesters opposed the construction of a copper smelting plant. This plant was a central part of the local government’s plan for economic rebuilding after the Sichuan earthquake, which had devastated the city’s economy in

2008 (Bradsher July 4, 2012). However, residents feared that the byproducts of the plant would create pollution that might have negative influences on their health and environment.

The protests attracted thousands of participants. They were mainly organized by students in Shifang, but expanded to include more citizens. Violent interactions reportedly took place between local police and the protesters. Some of the students were detained over the course of the protest. On July 3, after three days of protesting, the local government conceded, announcing the shut down of the copper plant (Blanchard July 8, 2012).

Qidong Pipeline Protests

The Qidong protests occurred on July 28, 2012. Thousands of protesters took to the streets of Qidong to protest a pipeline project that would take waste water from a Japanese-owned paper mill to the port near Qidong. Protesters stormed the local government office, holding the mayor captive. That same day, the local government cancelled construction on the project. The protests were largely organized and motivated by the fishing industry in Qidong, which feared that the pollution would have a negative impact on their livelihood. They were also nationalistic in nature because the plant was owned by a Japanese company (Lin July 30, 2012).

Censorship in both of these environmental protests focused on removing social media posts about the protest event, in particular posts from Shifang and Qidong. However, censorship in other geographical areas was somewhat delayed, either because the local government needed time to communicate with the central government about the protest event, or because of political competition between provinces. Thus censorship was not employed as consistently and quickly as in the Tibetan case, which provides a useful contrast. The variation over geography within these environmental protests allows me to measure the influence of local information accessibility on the spread of information, which I detail in the analysis section.

4.2.3 Larger Ideological Conflict: Anti-Japanese Protests

At the beginning of 2012, the mayor of Tokyo, Shintaro Ishihara announced his plans to nationalize the Diaoyu/Senkaku islands, a group of islands claimed by both China and Japan, but at the time owned privately by a Japanese businessman. While initially, this move was only part of Ishihara's election campaign, eventually the Japanese government moved to support the nationalization of the islands in July 2012 because of political pressure.

In reaction and in the lead up to the nationalization of the islands, huge protests erupted in China. Hundreds of cities witnessed anti-Japanese protests during August and September of 2012 (Weiss 2014). The protests occurred all over China, and included not only peaceful demonstrations, but also looting of Japanese business and Japanese-owned property.

While the protests were widespread, the Chinese government's reaction was understandably mixed. Since Japan did not back down from nationalization, the existence of anti-Japanese protests buttressed China's power in negotiations vis-a-vis Japan. The government could point to these protests as domestic pressure that would not allow it to accede to nationalization of the islands (Weiss 2013).

At the same time, these protests included citizens with grievances against the Chinese government itself. The protests were not purely against Japan, and some of the protesters expressed anti-government sentiment, reflecting grievances over other social injustices, or support for recently-ousted leader Bo Xilai, in direct defiance of government policy (Weiss 2014). Anti-government sentiment could be particularly destabilizing in the lead up to the 18th Party Congress, scheduled for November 2012.

Official policy reflected this mixed view on the protests. Historically, elites have used nationalist anti-Japan sentiment as a vehicle of voicing policy objections and advocating for their own position within government, and as a result, much conflict exists about the correct response to the

protests within the government (He 2007). The government finally did crackdown on September 18, 2012, when the largest and most violent protests took place. However, before then, the government reaction was mixed. Certain cities, especially large cities, cracked down on the protests, while smaller cities did not (Weiss 2014). Some websites heavily censored anti-Japanese sentiment, while others let it go relatively uncontrolled. This variation likely reflected splits within the government about how to deal with the protests and the conflict with Japan.

The anti-Japanese protests, therefore, present a case in which to study the influence of censorship when censorship policy is not consistent across websites because of splits within the government. Without uniform censorship across websites, discussion about the islands is very prevalent before the September 18th crackdown. However, despite non-uniform censorship, I show that content filtering reduces new postings about the Diaoyu Islands within websites so that even small-scale content filtering can reduce discussion within local communities.

4.3 Content Filtering Slows the Spread of Information: Evidence

In order to study the efficacy of censorship across each of these very different protest events, I collected social media postings related to the event during the time period that they occurred. I collected these social media postings before the Chinese government was able to censor them, and later returned to the URL to see whether they had been removed from the internet. Thus I measured which posts were censored for each of these events, on which websites and, in many cases, the location of the post itself.

I now take a closer look at each protest event, and using the variation created in censorship across websites, geography, and time, show how content filtering is effective in slowing the spread

of information and how, when government policy is inconsistent, the inconsistencies of government policy caused by disagreements and competition between localities undermines the efficacy of censorship.

4.3.1 More Discussion of Self-Immolations on the Weekends When Censorship is Lower

To study how information friction through content filtering influences the spread of news surrounding self-immolation events, I collected a random sample of social media postings related to self-immolations between March of 2011 and July of 2013 before the Chinese government was able to censor them. I then checked to see which of these posts had been removed from the internet.

Discussion of self-immolations on social media in China naturally clusters around self-immolation events. In this section, I define “bursts” of social media posts about an event as the spike in volume of social media discussion about an event at the time of the event. Social media is by nature bursty (Ratkiewicz et al. 2010), but some events receive more attention than others. In this context, since I obtain posts before censorship, I measure how many posts were *written* about the event. However, depending on how quickly censorship occurs, many of these posts were only *accessible* to readers for a short period of time. The government censors many of the postings about self-immolations and is largely successful – around 80% of all social media posts within a burst associated with a self-immolation event are censored.

Despite relatively uniform censorship across self-immolation events, some self-immolations have more posts than others – some have longer bursts, or more social media posts about that immolation event. Figure 4.1 shows the variation in bursts across the 120 self-immolation events between 2011 and 2013. Some events receive barely any attention, while others have a large amount of social media discussion associated with the event.

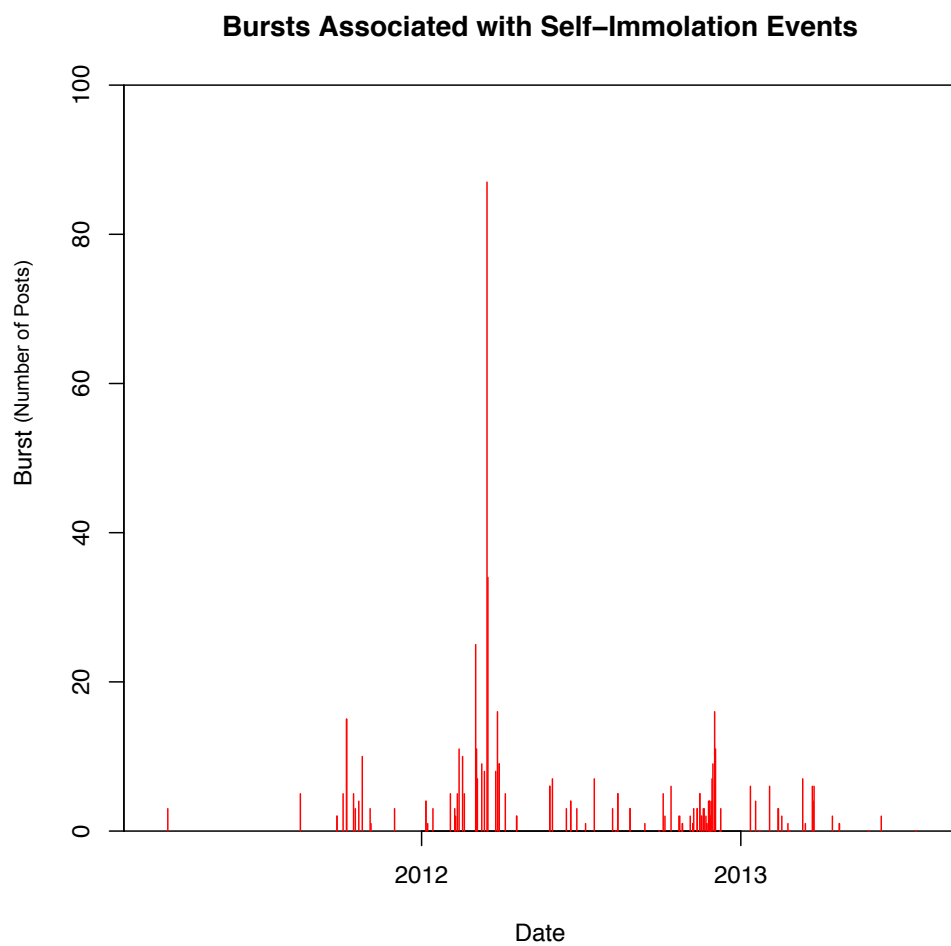


Figure 4.1: *Length of bursts for 120 self-immolation events between 2011 and 2013*

Why would some immolations receive more attention than others? It could be that the nature of the event was such that some self-immolations received more attention from the public than others. The age of the immolator is usually something people note when discussing self-immolation events, with younger immolators often discussed with more apparent grief than older immolators. Monks who self-immolate might have a larger network of followers, or be viewed more sympathetically than lay immolators, leading to more attention about the event. Self-immolation events that appear in clusters might build on each other, generating more attention.

However, in an environment of high censorship, the fact that some immolators receive barely any online attention at all and that the spread of information about self-immolations is overall so stifled suggests that the length of the burst could be explained by variation in friction caused by the control of information. If difference in burst length between self-immolation events were due to variation in information friction, we would expect that social media posts that were online for a longer period of time would also have more time to be re-shared by others. The faster censorship, the fewer people would know about the event, and the fewer posts would be written about it. The quicker that censors react to an event, the less online discussion about that event.

I do not have realtime data on censorship of self-immolation events because the infrastructure required to detect realtime censorship over such a long time period is prohibitively large. However, real time analyses of censorship over short time periods can be conducted, and a few authors have uncovered the regular schedule of censors. Using real-time data collected by King, Pan and Roberts (2013a), which were collected around the same time that the posts within my sample were written, I find that censorship is relatively low on the weekends relative to the weekdays. King, Pan and Roberts (2013a) find that censorship generally occurs within one day of posting. A post written Friday is most likely censored on Friday. However, a post written on Friday is second most likely to be censored on Monday, not on Saturday or Sunday. This suggests that fewer censors are working

on the weekends than during the weekdays.

Self-immolations, however, can happen on any day of the week, and do. An analysis of all self-immolation events over the past two years show that self-immolations are no more likely to happen on one day of the week or another. Since the act of self-immolation is so desperate, there is likely very little strategy involved in the particular day of the week chosen.

If information friction were effective in stopping the spread of information, we might expect that bursts related to self-immolations would be longer on the weekend, given that it might take longer for censors to locate and delete these postings and therefore provide a longer period of time for others to read and repost these social media posts. Indeed, bursts associated with self-immolations on the weekends are significantly longer than those associated with self-immolations that occur on a weekday.

Of course, the length of the social media burst could be due to other variables besides censorship. Monks might be more more likely to gain a larger following. The age of the self-immolator could be related to both the timing and the following. In order to control for these variables, I collected data about the specific circumstances of each self-immolation event, including the age, whether the self-immolator was a monk, and the time since the last self-immolation to capture any clustering effects.

I model the length of the burst using a negative binomial regression, where the length of the burst is the dependent variable and whether or not the self-immolation occurred on a weekend is the main independent variable of interest. Controlling of characteristics of the self-immolator, I find that whether the event was on a weekend is still a significant predictor of the length of the burst. Simulations from the model are included in Figure 4.2, showing a positive effect of an event happening on the weekend on the expected number of posts within the burst.

The finding that the speed of content filtering influences the number of posts about a self-

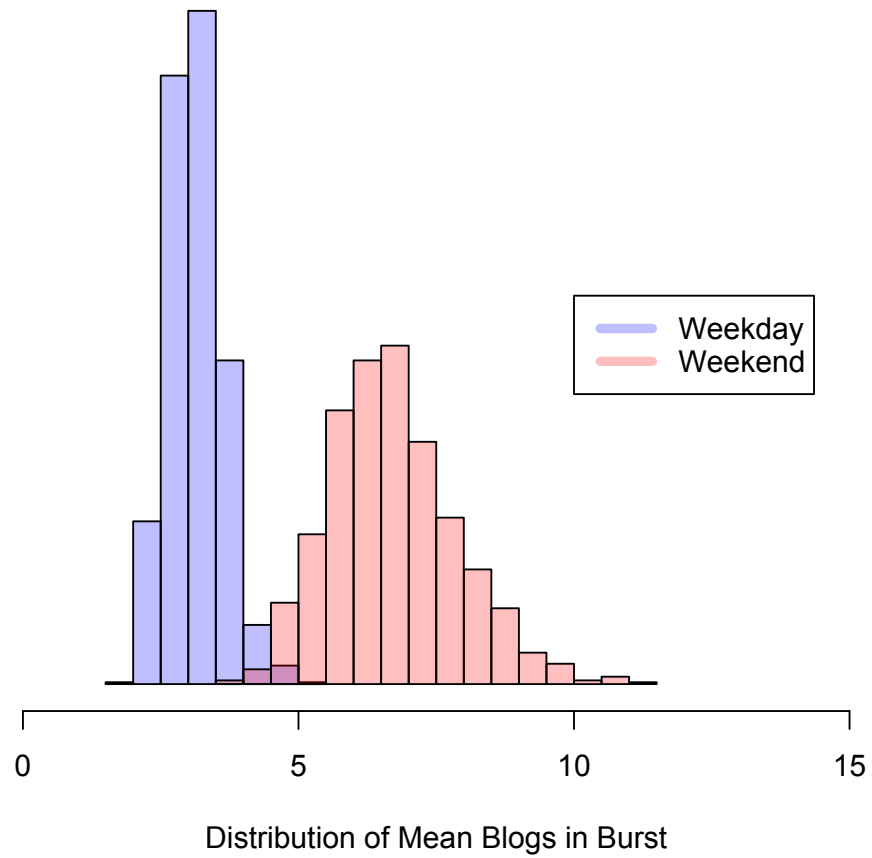


Figure 4.2: *Weekend self-immolations have longer bursts than weekday self-immolations, negative binomial model.*

immolation event indicates that small costs of access to information, such as the timing of censorship, influence the spread of information about protest events throughout China. Even in this case, where fear and self-censorship should be constant across immolation events, the timing of censorship dictated by censors' schedules influences the number of people who knew about an event.

This suggests that information friction, the decreased accessibility of information does indeed slow the spread of information in China. If censors are slightly faster at taking posts down for a self-immolation event, information about that event is slightly more costly to netizens than if they remove posts more slowly. Faster censorship means fewer people access that information and fewer write new posts about the event, further decreasing accessibility of information about the protest. I now turn to protests where the Chinese government is less consistent in censorship.

4.3.2 Environmental Protest: Local Censorship Influences Local Discussion

To study the influence of information friction on the spread of information about environmental protests in Shifang and Qidong, I collect geo-located social media posts from Sina Weibo relating to both protest events. These protest events were significant political events, and netizens all over China discussed them. However, censorship exhibited much variation across localities in China. Figure 4.3 shows censorship rates for both Shifang and Qidong in the Chinese provinces on the first and second day of each protest. Clearly, there is variation in censorship across locality; in particular, censorship initially focused on the location of the protest event itself and later spread to the entire country, but also some localities picked up censorship of the event more quickly than others.

If small costs of information were significant in reducing access to information during these protest events, we would expect that local censorship would influence the local spread of informa-

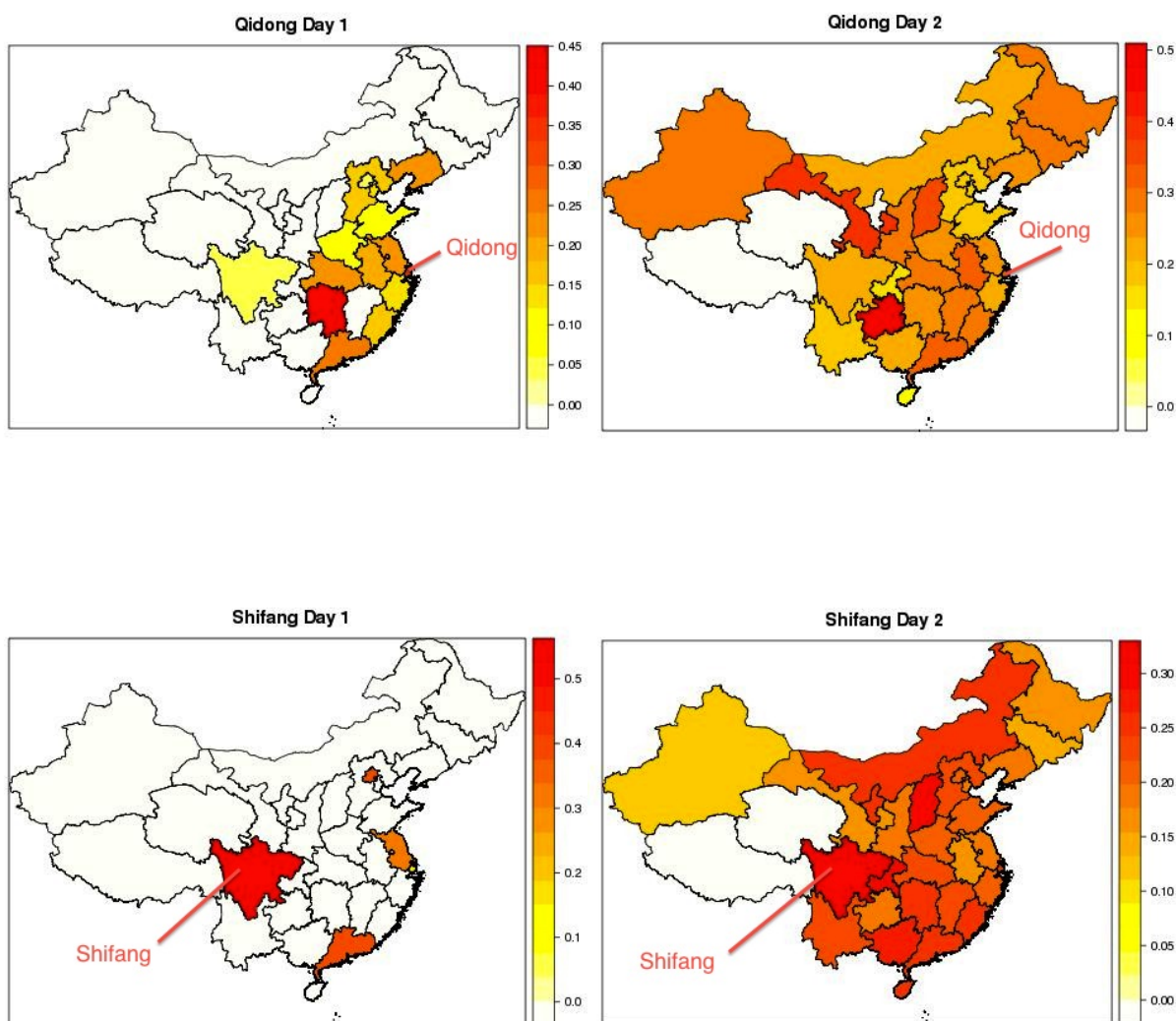


Figure 4.3: Geography of protest events for Qidong and Shifang environmental protests. Censorship begins with high variation across provinces but ends with low variation. White provinces indicate there is not enough data to provide censorship information.

tion. People tend to be more connected with others within their same locality. Even in an online world where sharing messages with people across the country is as easy as sharing messages with people next door, networks of friends that develop on the ground also mean that people who are closer to each other are more connected online. We would expect that if content filtering were higher in one locality, people in that locality would be less likely to come across information about the protest event; information about the protest would be less accessible for that locality. High content filtering in Guangxi and low content filtering in Beijing will mean that on average people in Guangxi province are less likely to read information about the protest because their friends' posts have been taken down, whereas talk about the protest in Beijing is more likely to go viral.

To measure the “local effect” of censorship, I gathered information on the total number of posts from each city during the duration of these two protest events and the number of these posts that were censored. If censorship has a local friction effect, we would expect that if locality A and locality B have the same number of posts on day 1, but locality A has a higher censorship rate, then locality A will have fewer new posts on Day 2. To estimate this, I ran a negative binomial regression of the number of posts in a locality on the number of posts in that locality the day before and the censorship rate the day before. To capture locality-specific and date-specific effects, I include time and city fixed effects. I combine the Shifang and Qidong protests, but also include an indicator variable for the protest event, in order to capture the fact that the two protests could have had inherently different levels of interest.

I find that the censorship rate for the previous day within the locality significantly decreases the number of new postings from that locality on the following day. Figure 4.4 shows simulated results from the model of the expected number of posts for a variety of censorship rates the previous day. The expected number of posts is decreasing with the censorship rate.

This “local effect” of censorship suggests that content filtering is still somewhat effective even

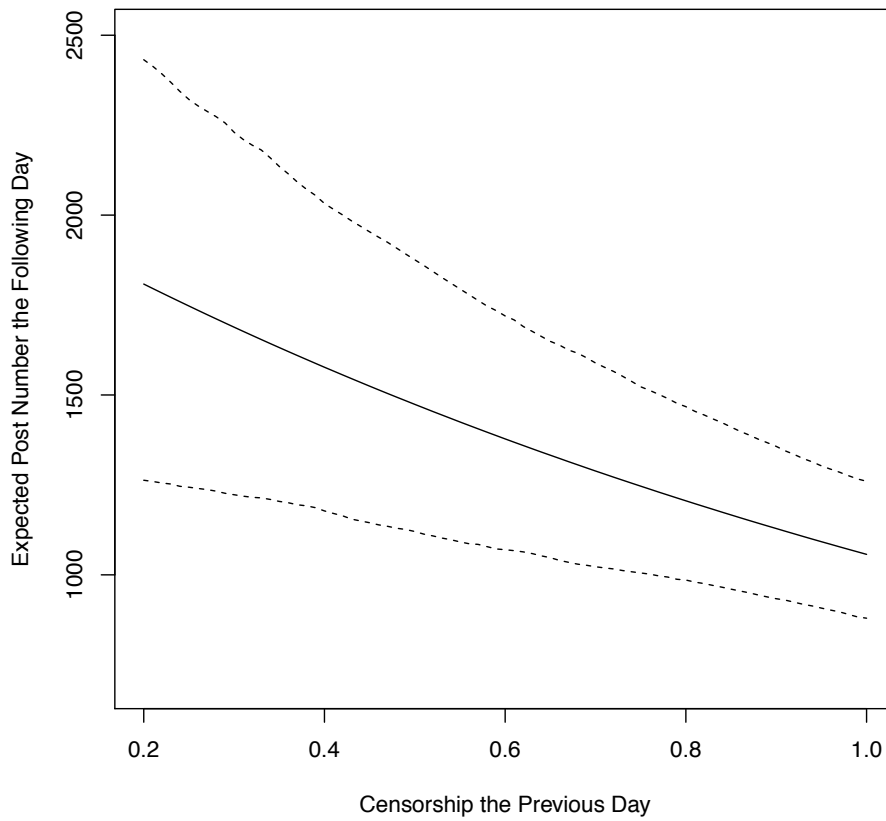


Figure 4.4: *Marginal effect of censorship on the number of posts in the following day, across cities in Qidong and Shifang environmental protests.*

when it is not completely consistent across websites. It also implies that more consistent content filtering across localities would be more effective in stifling discussion. This suggests that competition between localities, or miscommunications between groups would make friction less effective. I now turn to the last example, where censorship is more variable across websites and show content filtering, while much less effective overall, is still somewhat effective at reducing the number of new posts within websites.

4.3.3 Anti-Japan: Websites with Higher Censorship Have Less Discussion

To study the efficacy of censorship for anti-Japan protests, I collected thousands of social media posts from 12 different websites surrounding the protests against Japan in September of 2012. Protests erupted in over 200 cities during this time period. The largest protests occurred on September 18, 2012 which is “National Humiliation Day” in China, the day preceding the Japanese invasion of Manchuria in 1931. After September 18, the government began explicitly cracking down on the protests, and protests ceased after that time period.

The data from the anti-Japan protests is useful in measuring the effect of friction in the discussion of the protests on social media because there is a large variation in censorship of these posts across websites. While the Chinese government censorship program is focused on censoring information about protests, anti-Japanese protests are somewhat of a gray area and are occasionally state-coordinated. There is also evidence of significant sub-national variation in the extent to which anti-Japan protests are discouraged (Weiss 2014). This inconsistency in censorship is reflected in the websites, which are likely getting censorship directives from a variety of different sources within the government.

If the friction caused by censorship were effective in slowing the spread of information, we would expect higher levels of censorship within a website would be associated with fewer sub-

sequent posts on that website. There are two reasons this might happen – first, content filtering removes posts, making it seem like little discussion is happening on that website. When a netizen wants to discuss her opinion on the Diaoyu Islands crisis, then she might be less likely to go to a website with very few posts about anti-Japan because there is little existing discussion to add to. Second, frequent visitors to the site are less likely to read about the topic if it has been censored and may therefore be less likely to become interested in the topic and generate new posts. The fact that the posts are missing might go unnoticed but the lack of posts themselves will create frictions for further discussion.

On the other hand, if information friction caused by the selective removal of posts were not effective in dissuading discussion, netizens would create new postings on the website regardless of censorship on the website. People would access information about the protests regardless of what website they follow, and they would post their opinions on this no matter what website they blogged on. Partial censorship one day would not decrease postings the next because persistent users would find information about the crisis.

Using variation across websites with the anti-Japan data, I conduct two tests to see whether within-website censorship is effective in stifling discussion. First, I see whether websites with more censorship also have fewer postings during September 2012. This is true, consistent with effective censorship – websites with higher levels of censorship also have fewer posts during the time period, as showing in Figure 4.5.

This first test does not leverage the time variation of the postings – it simply looks for a correlation between censorship and the number of postings over the month of September. We know from previous research that the websites within China focus their censorship efforts within bursts – or when many people begin talking about the topic of interest all at the same time. If censorship were effective at stopping discussion on a website, we should expect that for social media websites that

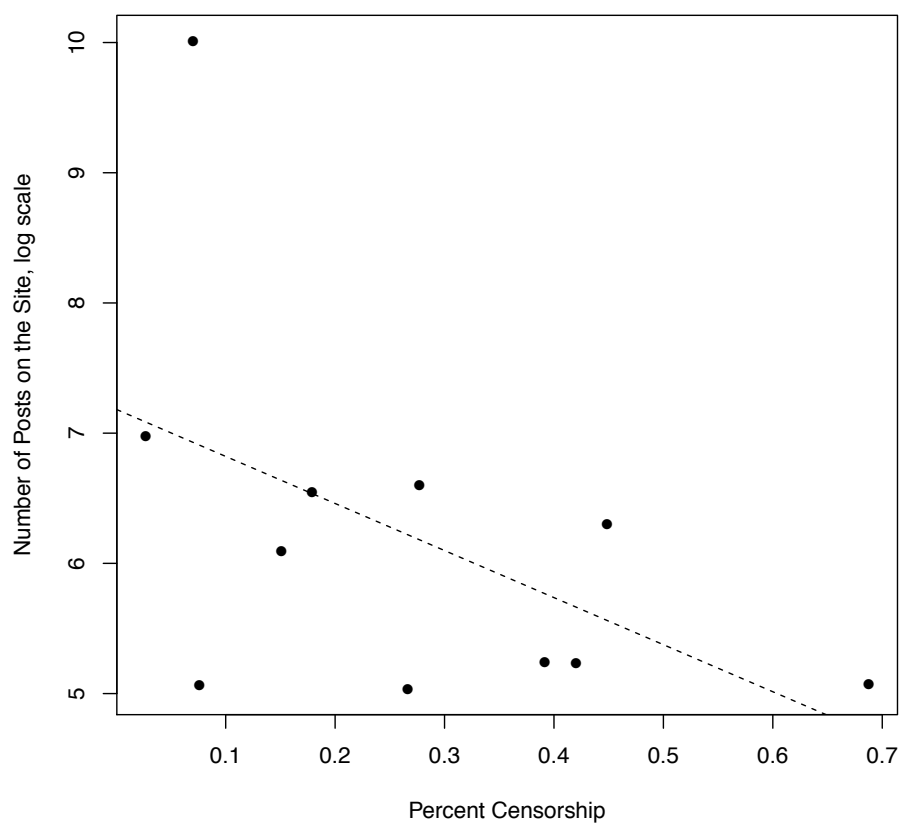


Figure 4.5: Websites with higher levels of censorship also have fewer posts, anti-Japan protests.

delete most postings within a burst, the burst itself would be shorter than those that do not censor posts within the burst.

For example, Figure 4.6 shows the number of posts on two similar websites in China. The first website has an overall high censorship level of discussion of the Diaoyu Islands crisis, with censorship concentrated in the burst. The second website has an overall lower censorship level and does not increase censorship within the burst. Consistent with the notion that censorship decreases the length of the burst, the first website has a much more peaked burst, while the second website has a much more natural increase and decrease in discussion of the topic over time.

To extend the analysis to all websites, I measure the length of the bursts for each of the 12 websites surrounding the September 18th protests. For some websites, the burst of conversation surrounding the September 18th protest lasts five days, for some websites, the burst is only for the day of September 18th. I then calculate the percent censorship during the burst. Consistent with the notion that higher censorship on a website decreases discussion on that website, the length of the burst and censorship within the burst are significantly negatively correlated. High censorship of the September 18th protest events also means fewer postings within websites and a shorter length of the burst on the website.

Similar to the previous example, since content filtering is effective within websites, the evidence suggests that information friction is more effective the more consistent information friction is across websites. Ideological disagreement and the lack of coordination across websites undermines the government's censorship policy. Unlike the case of the self-immolations, when all websites agreed on censorship and therefore very few posts about self-immolations could be located online, the total amount of discussion online about the anti-Japanese protests was much more widespread than about self-immolations.

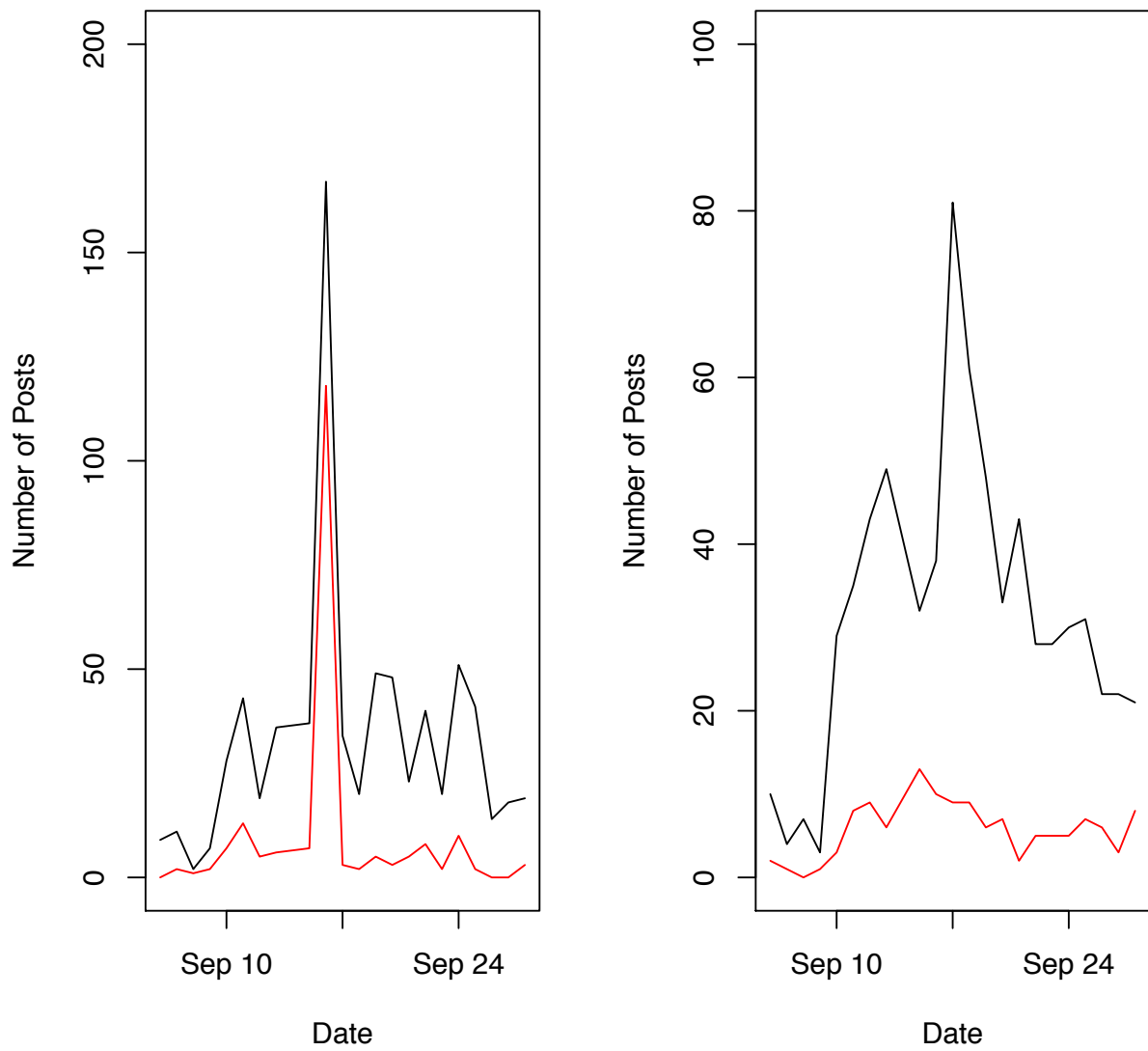


Figure 4.6: *Censorship profile during anti-japanese protests over two websites. Red line indicates censored post, black line indicates all posts.*

4.4 Conclusion

This chapter shows that despite the increased availability of information online, the Chinese government uses information friction in the form of content filtering to decrease the accessibility of information related to protest events, thereby reducing the spread of information about protests in China. The success of the content filtering program rests not keeping information secret or intimidating people to keep information secret, but on creating frictions by slightly decreasing access to information to make it unlikely that large numbers of people will become aware of the protests. This strategy is particularly effective when governments can organize a consistent policy of censorship across factions and localities.

Chapter 5

Newspaper Coordination and Its Internet Impact

In previous chapters, I show that fear is not the main mechanism through which governments and organized interests slow the spread of information. I show that *information friction*, or small and primarily invisible costs of access to information, is instead the primary mechanism through which censorship is effective in the digital age.

In this chapter, I move to a discussion of the ways in which governments and interest groups *promote* information in a the digital age and the mechanisms through which this promotion is effective. I argue that the promotion of information is still possible in an information age through a mechanism I call *information flooding*, or the coordinated effort to distribute information. Information flooding occurs when a group or government distributes its viewpoint by ensuring that a particular piece of information or a particular perspective is repeated from many different sources in the news media or social media. Because of the repetition, this information is highly accessible, if not impossible for citizens to avoid. Citizens are likely to come across the information and are likely to share it with others.

Information flooding is the opposite of information friction as it has a multiplier effect on the information environment. The more information that a group distributes initially, the more third parties will re-share it, or pick it up within their own writing. Flooding is a strategy only available to groups with concentrated interests like the government and interest groups because these groups have the resources and incentives to force third-parties to share the information.

To empirically test whether information flooding by the Chinese government influences the spread of information within the Chinese blogosphere, in this chapter I collect all articles published by 17 different provincial newspapers over the course of 2012. I identify “flooding topics” as particular topics that are coordinated across many of the 17 newspapers, and therefore were extremely low cost to the public. Despite the fact that many of these articles would be considered “propaganda”, I show that phrases in articles that are coordinated across more newspapers are much more likely to appear in other writing in both the domestic and international blogosphere than topics and wording that is not coordinated across provincial newspapers. This evidence suggests that flooding of Chinese propaganda influences the writing and wording of bloggers all over the world.

This chapter is outlined as follows. In the next section, I describe information flooding, and how it differs from previous theories of the efficacy of propaganda. In the following section, I discuss the use of flooding strategies by the Chinese government. I then describe the data I use to test this theory. In the last section, I describe the results and estimates from my test on how information flooding influences the spread of information online.

5.1 Information Flooding and Propaganda

Using similar logic to scholars who argue that censorship is impossible in a digital age, many scholars have argued that propaganda, or the promotion of information, is also outdated in the age

of the internet (Lynch 1999; Lieberthal 1995). These scholars maintain the propaganda can only be effective when the state can control the agenda, which occurs when the number of sources of information is constrained. As the number of media sources has proliferated with the advent of the internet, consumers of information have more choices over the sources of information. These authors argue that consumers of information select out of state media sources and into sources with more “reliable” information (Stockmann 2012).

A few authors oppose this view, arguing that while state media is recognizable, propaganda is a signal of government power and therefore is closely followed by citizens (Huang and Li 2013). Propaganda, in these authors’ views, creates norms that citizens are trained to follow (Brady 2009). Even if this propaganda is unbelievable, by inducing participation in propaganda, the state can create rituals and standards that encourage compliance (Wedeen 1999). In particular, in China, the government emphasizes propaganda to promote cultural governance, appealing to citizens emotions to prevent protests and keep them in line with government policy (Perry 2013).

In this chapter, I show that propaganda can be effective simply through the prevalence of information. Governments and organized interest groups coordinate propaganda by repeating information from multiple sources so that it is low-cost to citizens. Because such repetition makes information highly available to citizens, the population will be more likely to consume propaganda, regardless of whether they can identify the source. The “flooding” of information in the news media and blogosphere by governments and interest groups works not so much to signal power as to prioritize the consumption of government-produced news over news produced by other groups or by citizens themselves. Propaganda is effective because governments and interest groups have the resources to make it easy to access and low cost, and low-cost stories are more likely to reverberate throughout the blogosphere.

Coordination of propaganda has long been used by governments and other organized groups

to promote information. China's 1977 Propaganda Directive explicitly directs the Propaganda Department to coordinate stories between the news media in order "the promote the CCP's current line" (Brady 2009). In the United States, the Republican party is known for coordinating messages across newspapers and television, even to the extent that right-wing media outlets will use identical wordings to describe the same news point.

More recently, governments and mobilized interests around the world have organized "internet armies" in order to flood the blogosphere with particular perspectives at a coordinated moment in time. Notoriously, the Chinese "Fifty Cent Party" allegedly pays Chinese netizens 50 cents each for Party-sanctioned posts. While the Chinese have been criticized for this strategy, other governments have adopted similar strategies, including Israel (Liphshiz January 19, 2009), which pays its students to write pro-Israel comments on Facebook and Twitter, and recently Turkey, which has a 6,000 member social media team to leave pro-government posts (Albayrak and Parkinson September 16, 2013).

Chinese government officials and companies also use coordinated flooding strategies to bolster support. These groups will pay public relations companies and sometimes even regular citizens to post positive accounts of them online, similar to companies in the U.S., who will sometimes pay people to write positive reviews of their company on the web. Such actions have caused a certain amount of scandal in China, both because these officials and companies sometimes pay newspapers for articles that reflect positively on them, and because they have sometimes been successful in paying internet content providers to censor negative content about them (Chen et al. February 19, 2013).

If governments, companies, and politicians were using such messages to signal their own strength, as some authors have argued, they would want to take credit for these online messages. However, governments typically try to cover up the fact that they pay people to write

pro-government messages. Instead, they prefer that it appears as if “everyone” is writing pro-government comments or reporting a news story the government finds favorable. Part of the strategy of information flooding is issuing propaganda from many different sources, so as to disguise the fact that the information originated with the government.

By creating a multiplier effect in the news media and online, information flooding if successful can be worth the investment. The more sources a government or interest group can pay or force to cover their story from its perspective, the more other news groups and other social media users pick that story up and share it with others. Flooding begets more flooding, and if effectively done, this domino effect of information dissemination can be exponential. What began as a propaganda message, can seem like a online event created by citizens, as more and more people read and share the story.

Flooding can also create friction for stories that are less desirable to the government. In interviews with Fifty Cent Party members, Ai Weiwei reveals that Fifty Cent Party members are often instructed to distract from current stories that are less desirable to the government (“An insider’s account of the 50 Cent Party.” May 12, 2011*b*). Flooding of entertainment and “soft news” stories brings these stories to the forefront at the expense of stories of protests of government wrongdoings, for example. If citizens are distracted by the accessibility of flooded stories, they are less likely to read other stories.

5.1.1 Intra-Party Conflict and Coordination

Though the Chinese government has an extensive apparatus for coordinating newspaper stories across newspapers, when conflict exists within the government, such coordination is much more difficult to attain. Shih (2008*a*) shows that provincial newspapers mention Jiang Zemin’s slogan “Three Represents” at different rates, depending on their own political loyalty. We should expect

that even when the government wants to use flooding, if there are conflicts within the government, the amount of coordination between newspapers will reflect those conflicts. Similar to information friction in the last chapter, as more competition and divisions emerge between different bureaucracies or government groups, information flooding will be undermined by dissenters.

5.2 Information Flooding in China

Coordination of information to produce such flooding is key to the information strategies of the Chinese Propaganda system. The Chinese government is in the perfect position to coordinate because it has the resources and infrastructure to do so. First, the institution of propaganda in China is built in a way that makes coordination easy. The Propaganda Department is the most extensive bureaucracy in the Chinese Communist Party, infiltrating every level of government (Lieberthal 1995). It is managed and led directly from the top, within the top Party organ, the Politburo.

From the very locus of power within the Party, messages are coordinated throughout the news media within China, through every medium of news, including television, print media and radio.¹ The government controls the personnel in every major media organization within China, and requires each journalist to be government-certified. For day-to-day monitoring of content, the government issues propaganda directives to editors who then decide what is included in the newspaper (Zhao 2008). Post-publication monitoring is conducted by retired propaganda officials who make sure that newspapers are following the issued guidelines (Brady 2009).

The extent of newspaper coordination within China has waxed and waned throughout recent Chinese history. During the Maoist period and Cultural Revolution, articles within the People's Daily coordinated news around the country – smaller newspapers would reprint People's Daily articles when instructed (Yu 1964). With reform and opening after 1979, the coordination of

¹In this chapter, I focus explicitly on coordination within the print news media in China.

news within China was significantly loosened and the Central Propaganda Department was weakened (Brady 2009). In the lead up to the Tiananmen Square pro-democracy movement in 1989, newspapers were less coordinated, and newspapers such as the *World Economic Herald* became well-known as critics of the Party.

The events of 1989 caused a complete reversal in the CCP's strategy toward propaganda and coordination. After the crisis in Tiananmen Square, the government decided to re-strengthen its grip on propaganda. For example, in 1990, one of the Party's leading news agencies *Xinhua* was close to bankrupt. However, the government decided to use *Xinhua* as the coordinating agency following 1989, instructing newspapers to follow *Xinhua*'s lead on important events and international news, much as they had done with the *People's Daily* during the 1960s (Brady 2009). *Xinhua* is now one of the most profitable newspapers in the country because it leads the coordination of news.

Though not as extensive, there is also some level of coordination on the blogosphere. As mentioned earlier, the Fifty Cent Party promotes government-sanctioned news online. In addition, the government is known to contact high-profile bloggers and important online opinion leaders before important events in order to coordinate political messages among highly-followed bloggers in China (National Academy for Propaganda Cadres 2011).

5.3 The Influence of Flooding on the Spread of Information

Having described the logic behind flooding, I now move to show that simply the prevalence of information, rather than the source of the information, has huge secondary influences on the spread of information throughout the blogosphere. If current theories of the influence of propaganda on the blogosphere were true, we would expect coordinated efforts at promoting information would not influence the spread of information because 1) citizens would not select into these official

sources and 2) citizens would recognize that this was propaganda, and would decide to share more “objective” information. If propaganda only produced fear in citizens or created norms for citizens to follow, there would be no reason for them to continue to share this information on the blogosphere, since they would have already identified it as propaganda.

If, however, information flooding were sufficient to encourage the spread of information, we would expect that highly coordinated news articles would reverberate around the blogosphere, even if the information were well-known to be associated with propaganda. To test this, I identify propaganda themes within a unique dataset of 17 provincial newspapers. I find that not just the topic of the propaganda theme, but the *particular wording of the article* spreads throughout both the domestic and also, more surprisingly, the international blogospheres.

5.3.1 Data

To test this theory, I collected every newspaper article from 17 provincial newspapers in China for the year 2012.² The newspapers were scraped from each newspaper’s “digital” website. These sites are different from the commercial news sites in that they contain only digital copies of the printed newspaper and do not include online advertisements or other extraneous information printed on commercial news websites. While some Chinese newspapers’ commercial websites differ from their printed papers, the articles on the digital website reflect the content of the printed newspaper exactly. I recorded the date on which each article was written, and other metadata information when available, including the author and images associated with the article.

I removed articles that did not contain words (some articles just contain pictures). In total, there are 82,972 newspaper articles with words within in my corpus. Most newspapers contain a little over 5,000 newspaper articles for the year. Smaller provinces, however, contain fewer articles –

²The newspapers covered Beijing, Chongqing, Fujian, Gansu, Guangdong, Guangxi, Hainan, Heilongjiang, Henan, Hunan, Jiangxi, Jilin, Inner Mongolia, Qinghai, Tibet, Yunnan, and Zhejiang.

Gansu and Inner Mongolia have about 4,000 articles each, whereas Hainan and Beijing have over 6,000.

5.3.2 Identification of Propaganda Themes

In order to understand how information flooding affects the spread of information, I first identify instances of propaganda and flooding within my data. To do this, I subsetting the newspapers to those that referenced the *Xinhua* newswire, which was about 1/8 of the articles within the corpus. As described above, newspapers typically reprint *Xinhua* at the direction of the central government. Therefore, the topics and themes covered within provincial newspaper articles citing *Xinhua* are likely to be representative of topics and themes the Chinese government intends to promote.

I estimated the Structural Topic Model (STM) Roberts, Stewart and Airolidi (2013) on the *Xinhua* subset to identify promoted “topics” that were commonly discussed during 2012. Topic modeling, rather than supervised learning methods, is an appropriate method for finding propaganda topics because I do not have to assume topics beforehand, but rather use the model to learn which topics are most common within the data. The Structural Topic Model, in particular, is appropriate because topics within this data are highly time-dependent. STM allows for the inclusion of covariates to better estimate topics, and I include time as a covariate to better estimate highly time-dependent topics. Topics in which newspapers reprinted or cited *Xinhua* I term “propaganda themes” for this time period.

The model identified 28 different propaganda themes within the data, shown in Figure 5.1. These themes belong to three main categories. First, there are themes that we would typically think of as propaganda. This includes a topic related to the promotion of “socialist harmonious society” and “scientific development outlook” (both ideological formulations of Hu Jintao), a topic related to rooting out corruption that peaks in March of 2012 with the downfall of Bo Xilai, and a topic

related to the power transition which peaks in November of 2012.

A closer reading of documents within these topics shows that these indeed read like propaganda articles. For example, a group of articles on May 29, 2012 reported a Politburo meeting that was held the previous day. Hu Jintao presided over the meeting, and the main point of the meeting was to deepen scientific, balanced development. This document was highly representative of the “scientific development outlook” topic. In another example, on March 11, 2012, only four days before Bo Xilai was removed from his post as mayor of Chongqing on corruption charges, provincial newspapers reported on a meeting at the National People’s Congress in which leaders of the Party stressed anti-corruption measures against Party officials. This article was representative of the corruption topic.

The second theme relates to policies implemented by the central government. These include the promotion of “green” policies and policies that protect the environment, policies related to the courts and reform of the courts, policies related to education (in particular to migrant children taking the national university entrance exam), and regulations related to food safety. Another example is a spike in articles from *Xinhua* relating to inflation in January of 2012, right after a sudden increase in inflation. A closer reading of these articles shows that the articles reassure readers that inflation is not out of hand and is expected to decrease. For example, a report on January 19 in several provincial newspapers reassures readers that based on the National Bureau of Statistics housing prices were falling or flat in most cities.

The third theme relates to international events where provincial newspapers reprint the *Xinhua* articles. Some of these are typical international events, like the Olympics in June of 2012 or the civil war in Syria, and some are very specific to China’s immediate international situation, such as the Diaoyu Islands crisis that occurred in the summer of 2012. These may not seem like propaganda themes a priori; however, it is well-known that China controls international news

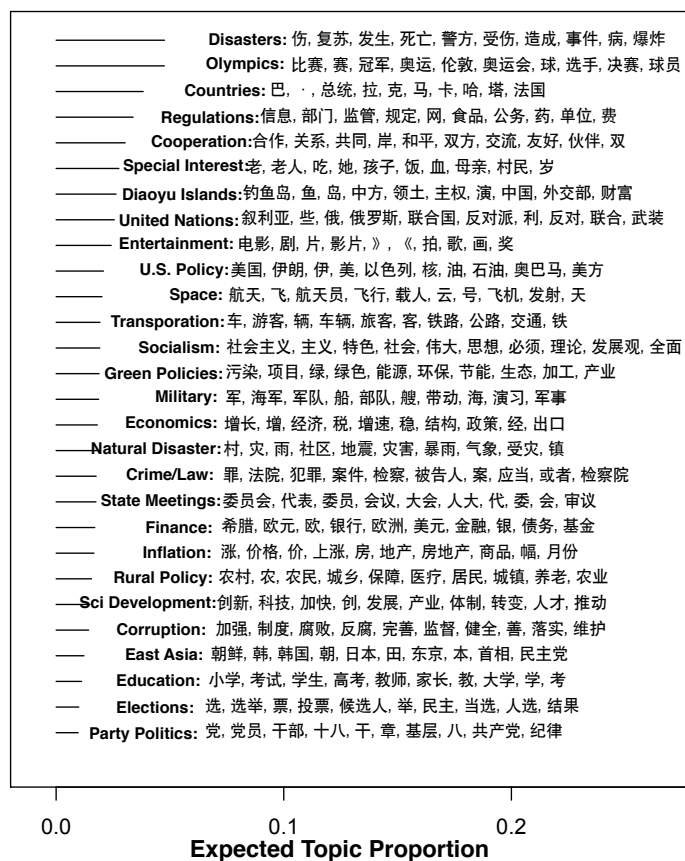


Figure 5.1: Propaganda themes identified with the Structural Topic Model.

through *Xinhua* in order to frame the population's perception of China's power in the international sphere, particularly on issues related to the United States or the contested areas within the South China Sea (Brady 2009).

For example, many provincial newspapers reprinted *Xinhua* articles about the Diaoyu Islands in the summer and fall of 2012, when large protests broke out across China about the purchase of the Diaoyu Islands by the Japanese government. These articles are often reports of Chinese leaders reiterating that the Diaoyu Islands are Chinese territory and that China would not tolerate any encroachment on them.

5.3.3 Selecting Representative Articles and Measuring Coordination

The Structural Topic Model estimates the proportion of each document within each topic. For each propaganda theme, I selected the 50 documents with the highest proportion of that theme. Since in total there were 1,450 articles representing propaganda themes, I couldn't read all of the articles, but I read a random sample to ensure that the model was selecting articles that were consistent with the themes described above.

Next, I estimated how coordinated each of these articles was across newspapers. To do this, for each article, I iterated through the entire corpus to estimate the similarity between that article and each of the other documents within the corpus. While many methods could be used to estimate similarity between documents, I opted for simplicity by choosing a method that counts the percentage of words each document contains of the other document. For the article of interest, A_B with words W_B and comparison article A_C with words W_C , I defined similarity as the geometric mean between the percentage of words in the article of interest that appear in the comparison article and the percentage of words within the comparison article that appear in the article of interest, formally written below:

$$S_{B,C} = \frac{\sum_1^{W_B} I(w_B \in w_C)}{2W_B} \frac{\sum_1^{W_C} I(w_C \in w_B)}{2W_C}$$

$S_{B,c}$ is typically quite low, about .1 for most articles. I trained the data based on a selection of documents that I knew were close copies of each other, and based on the distribution of the metric within these documents, selected a threshold of .8, counting two articles with $S_{B,C} > .8$ as coordinated. Therefore, for each selected article, I had a measure of how many newspapers had printed the same or extremely similar stories to that document.

5.3.4 Traditional Propaganda Articles are Most Coordinated

As shown in Figure 5.2, the most coordinated themes are those closest to traditional propaganda. The three themes with most coordination are scientific development outlook, socialist harmonious society, and corruption. On average, these themes were coordinated across 10 newspapers. The least coordinated themes were those related to non-political subjects, in particular special interest stories, entertainment and sports. For example, on average entertainment was coordinated with less than one other newspaper.

The fact that coordination aligns with our a priori ideas of propaganda is reassurance that the coordination measure reflects propaganda themes, rather than simply big news stories that occurred during the time period under consideration. While newspapers have some leeway over what news to report, it seems that they are at times strongly encouraged to reprint propaganda themes, which is reflected in the coordination across papers. Such coordination likely not only occurs among provincial newspapers, but also among private newspapers and national newspapers and through other media as well.

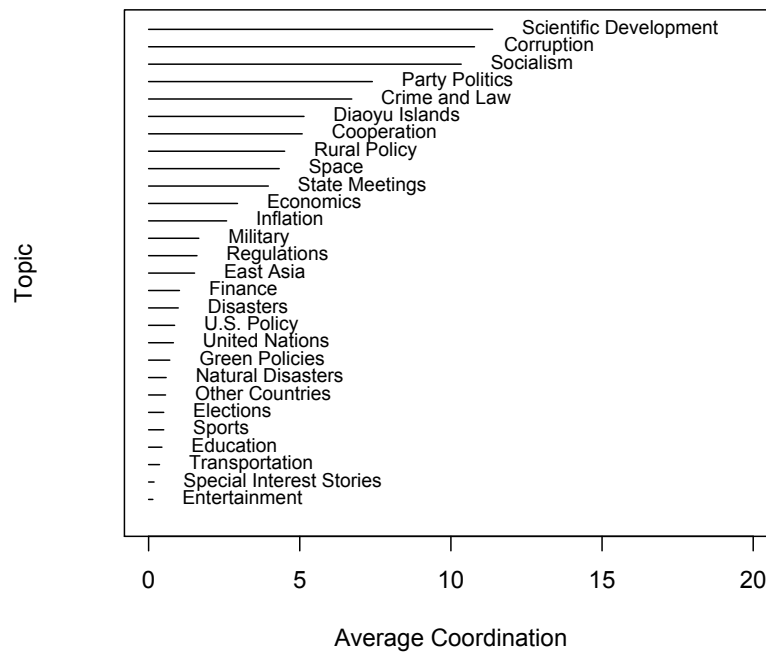


Figure 5.2: Average level of coordination for most frequent topics.

However, notice that no topic is perfectly coordinated across all newspapers. This reflects the fact that the Chinese government's coordination efforts are imperfect. Despite efforts to coordinate, some provincial newspapers likely defect, likely because the local government is associated with a different faction, or has different goals than the central government (Shih 2008a). For example, Jilin province is highly coordinated with other papers on the scientific development outlook, but exhibits much lower coordination than other papers on the Diaoyu Islands topic. On the other hand, Chongqing has high coordination on the Diaoyu Islands topic and relatively low coordination with scientific development outlook. Conflict between provinces and the provincial and central government undermines its flooding strategy.

5.3.5 How Does Coordination Influence the Spread of Information?

How does such coordination influence the spread of information? If information is sufficiently coordinated, are others more likely to reprint these stories? Do bloggers and other commentators pick up the same language used in the coordinated newspaper articles?

To answer these questions, for each of the 1,450 selected articles that represented the propaganda themes, I randomly selected ten fifteen-character strings. I then used automated methods to search each of these 14,500 strings on Google and record the number of search results returned.

I obtained three different search result metrics. First, I counted the number of search results Google returned overall. I also counted specifically the number of search results returned on sina.com.cn (using site:sina.com.cn within the search results), the most popular blogging site in China. Last, I counted the number of search results returned on blogspot.com, the most popular blogging site within the U.S., but which is blocked by the Great Firewall in China.³

³I used the Google API to do this, and upon close examination, the number of returned search results is fairly accurate. The number of search results when you simply search Google from a desktop is often very inaccurate, which is why using the API is important.

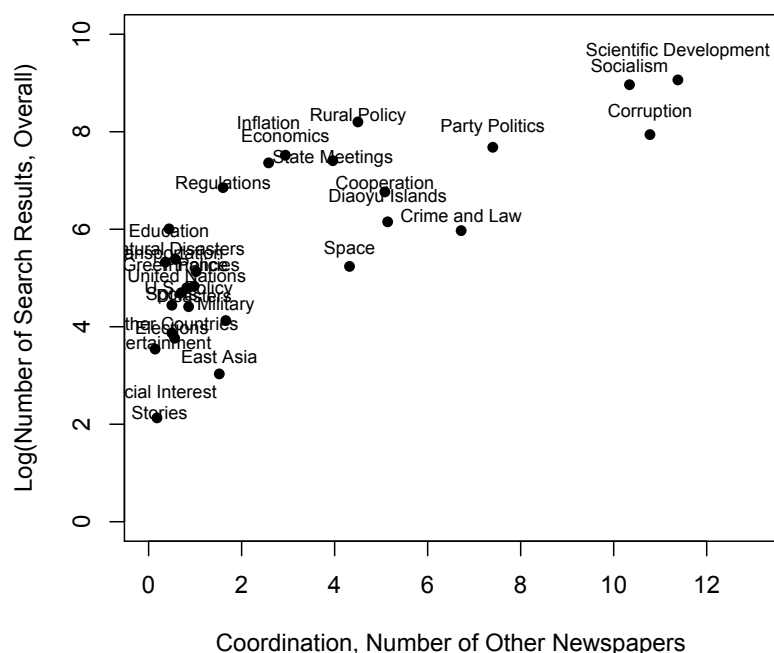


Figure 5.3: Relationship between average coordination and number of total Google search results, by topic.

I then measured the relationship between coordination and the number of search results, both within and outside of China. Figures 5.3, 5.4 and 5.5 shows the mean coordination by Topic on the x-axis, and the log of the search results on the y-axis. There is a very strong correlation between the degree of coordination and number of search results containing the strings of the coordinated articles. The correlation is particularly strong from the sina.com.cn search results, and less strong for the blogspot.com search results. This is preliminary evidence that coordination of Chinese propaganda reverberates throughout the blogosphere.

This is only preliminary evidence because it could be that across topics there is a strong correlation between coordination, but not within topics. To measure this, I estimated a linear regression

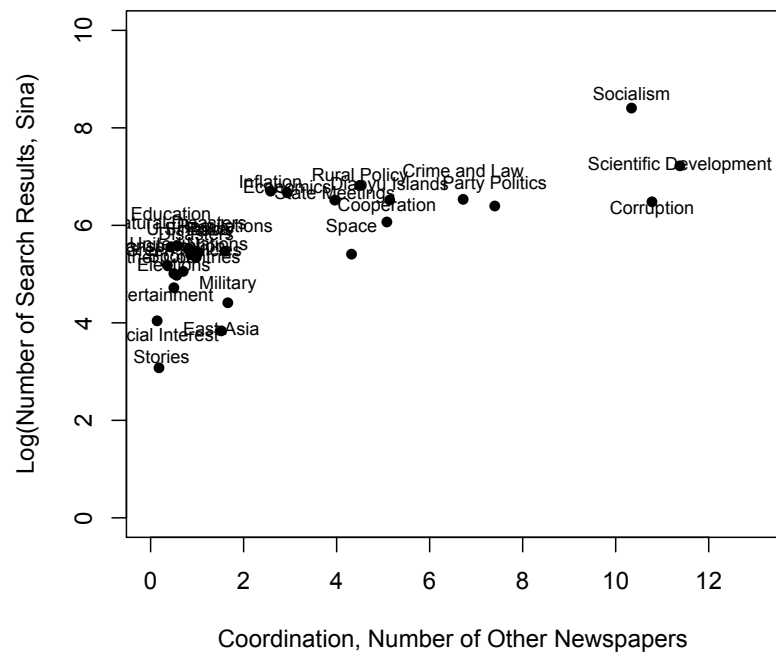


Figure 5.4: Relationship between average coordination and number of Google Search results on the site *sina.com.cn*, by topic,

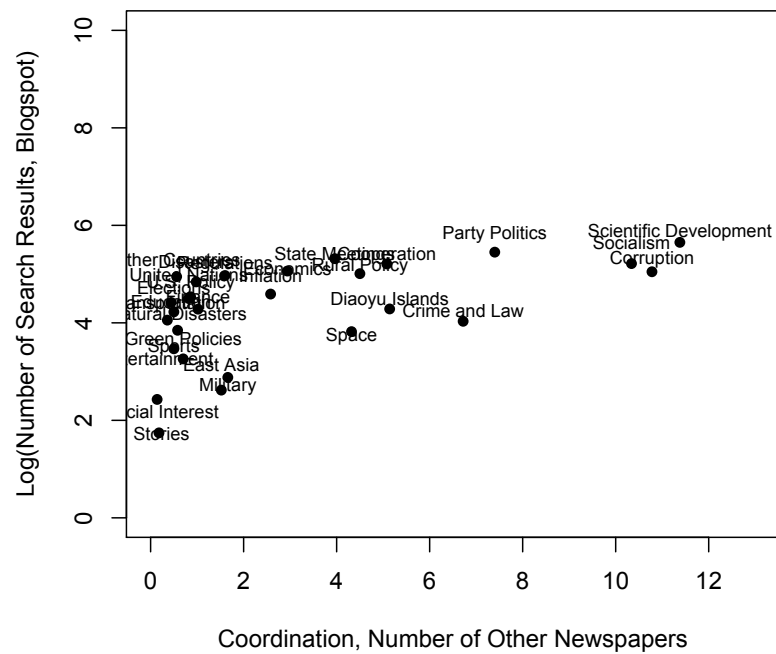


Figure 5.5: Relationship between average coordination and number of Google search results on the site *blogspot.com*, by topic.

of the search results for each document on the amount of coordination for each document, with a fixed effect for each topic.⁴

I ran a regression for each outcome, search overall, search for blogs on sina.com.cn, and search for blogs on blogspot.com. For each regression, the level of coordination is positively and significantly correlated with the number of search results. In Figure 5.6 I show the expected increase in number of search results given that the number of newspapers that coordinate increases from 10 to 11. As expected, the influence of coordination on sina and all search results is stronger than for blogspot. Increases by one newspaper increases the expected number of search results by 21 results on blogspot, 83 on sina, and 45 results overall.

It could be that these results are simply a reflection of coordination by Fifty Cent Party members who repost the same wording as in the provincial newspapers at the direction of the Chinese government. Certainly, this might explain why results for blogs sina.com.cn are much stronger than the search overall. However, we would not expect the Fifty Cent Party to write blogposts on websites outside of the Great Firewall, such as blogspot.com, which are less likely to be frequented by the average Chinese citizen. Thus, the evidence suggests that the link between coordination and blogging is not simply through other coordinated measures initiated by the CCP. Instead, the multiplicative relationship between coordination and search results likely indicates that the propaganda is being reprinted by regular Chinese citizens, and also citizens abroad.

The results also indicate that information spreads even if it is recognizably government-initiated propaganda. Despite the fact that the most coordinated topics are also the most obviously propaganda-related topics, there is a positive relationship between coordination and spread of information. This finding suggests that propaganda when paired with a flooding strategy can become commonly picked up throughout the blogosphere.

⁴I also estimated a hierarchical model, with a random intercept and random slope by topic. The results are the same as described here.

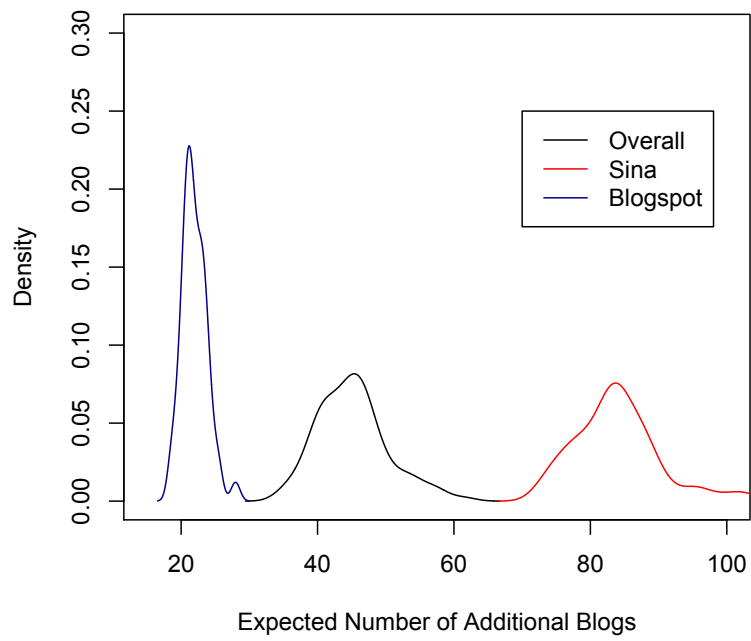


Figure 5.6: *Expected increase in search results if coordination increases from 10 to 11 newspapers.*

If information flooding is the mechanism through which the promotion of information can be successful, organized groups and governments have the advantage in using this strategy to promote information that they find favorable. A high level of coordination is required to issue the same information from many different sources, and such coordination is extremely costly since it involves not only communicating to these sources that they should disseminate a piece of information but also persuading these sources to disseminate the information. Groups with highly concentrated interests, such as governments and interest groups, are willing to pay the costs necessary to pursue information flooding strategy because the benefits they receive are concentrated. Like friction, which can disproportionately be used by organized groups, flooding too, is best wielded by the organized.

Chapter 6

Conclusion: Implications for the Information Age

In this manuscript, I seek to understand methods of online information control and the mechanisms through which they are effective. I find that intimidation and fear are not particularly effective in stalling the spread of information. Instead, these methods backfire to undermine the reputations and objectives of those implementing them. I demonstrate that less observable and more incremental methods of control are most effective in stalling the spread of information online. While consumers of information and the media will read and share information that is already mostly accessible and aggregated, governments and interest groups have the resources and incentives to decide whether to make information accessible, or instead to make it high cost to the public.

Organized groups, including the state, use strategies I call “information friction” and “information flooding” to control the spread of information. When information sheds a favorable light on a group, it will use its resources to coordinate information across media outlets, blogs, and websites, flooding the information environment so that the information is impossible to avoid. When, on the other hand, the information does not serve its purposes, it will use its resources to make that

information more inaccessible or disaggregated, frustrating users who access it. Consumers and the media have little time to search out such information, so the high costs of information reduces the spread of information to large numbers of people.

I test the efficacy of fear, friction, and flooding using the Chinese government's information control strategies as a case. I find experimentally with consumers of blogposts and observationally with bloggers that the awareness of censorship does not itself cause netizens to change topics or write fewer criticisms of the government. Instead, the observation of censorship piques users' interests in the topic, and they subsequently search for more information on the topic.

However, small variations in the extent of censorship during protest events caused by timing (in the case of Tibet), geography (in the case of local environmental protests), and websites (in the case of anti-Japan demonstrations) have huge influences on the spread of information within that time, locality, and website. Largely unobservable censorship, which makes information slightly less accessible, can have a significant influence on the spread of information.

Similarly, I find that the more coordinated the flooding strategy of the Chinese government across newspapers, the more the particular stories themselves reverberate throughout the blogosphere. The government can make information impossible to avoid and thus increases the spread of information online. Interestingly, the propaganda strategy of the Chinese government even influences the international blogosphere; bloggers on websites that are blocked in China are also more likely to use phrases that originated in coordinated Chinese newspapers.

When are friction and flooding strategies more or less useful? I claim that two factors determine how well groups can control the spread of information. First, competition between groups makes flooding and friction strategies more difficult to implement. As I show in this manuscript, when internal government conflicts over friction and flooding exist, the government has more difficulty controlling the spread of information. Second, the degree to which information is excludable,

or specific to that group, matters in how quickly the information spread. As a show in Chapter 4, information about Tibetan self-immolations is easier to contain than larger-scale environmental protests not only because of decreased competition between localities, but also because fewer people observe the self-immolations and therefore the Chinese police can more easily control the spread of information.

6.1 Implications for Autocracies and Democracies

Previous literature implied that in the digital age, information had to be kept completely secret, or would be easily spread to millions of people (Latham 2007; Besley and Prat 2006). It thus implied that either authoritarian governments would cave to public pressure and democratization with the advent of the internet (Ada et al. 2012; Bellin 2012; Ferdinand 2000), or that the government would need to use its monopoly on force and intimidation to create norms such that people wouldn't share incriminating information.

My findings imply that methods of control do not require that information is kept completely secret. Instead, small barriers to access can be extremely effective in keeping the number of people who know about that information to a minimum. Similarly, small increases in accessibility can spread information governments and interest groups would like to see promoted, distracting from other concurrent newsworthy stories.

That information friction and flooding, not intimidation, is the main force of online information control implies that not only governments, but any entity with sufficient interest in public opinion to be willing to pay resources to control information, such as organized groups, non-government organizations, political parties, and companies, will use similar strategies to control the spread of information. These strategies are also commonly used in democracies, where the independent media's revenues are being squeezed and therefore the media themselves will report stories where

there is easy access to information. While previous work has championed “e-democracy”, or the power of the masses in the information age (Dahlberg and Siapera 2007; Trippi 2005; Tolbert and McNeal 2003), my work suggests we should still be concerned about powerful groups’ monopolies on information, as they will be able to determine what information is accessible to consumers and reported by the media.

In future research, I will extend these results to information control in the United States, considered by many to be one of the countries with most unrestricted information. Like China, fear is probably not a primary method of information control in the U.S. Despite increasing surveillance online, the U.S. has a less significant history and more protective laws for consumers and producers of information. If citizens in China are undeterred by the observation of censorship, U.S. citizens should be equally as impervious and might be more likely to be curious about the U.S. government’s control of the web.

As a result, if information control is successful in democracies, it will occur through similar mechanisms outlined in this manuscript: information flooding and friction. For information flooding, we would expect to see organized groups collecting and distributing information they find favorable to the media and the public to make this information more accessible to consumers. For example, messages coordinated by political parties are likely repeated on the blogosphere in similar ways that domestic and international bloggers repeat messages coordinated by the Chinese Communist Party. Research sponsored by organized interests likely is repeated by media sources and around the web, eventually losing its attribution to the organized interest who supported it. This type of flooding in turn influences public opinion and participation.

Similarly, groups in the U.S. probably pursue strategies of information frictions similarly to local and central governments in China. U.S. companies take advantage of their preferential access of information about their products to hide product defects and create frictions by sealing court

cases that prevent the media access to this information (Efstathiou and Drajem June 6, 2013). The U.S. government sometimes engages in censorship efforts similar to that of the Chinese governments, encouraging media companies not to report on sensitive national security issues, such as those pertaining to Edward Snowden, or outlawing pictures of soldiers' caskets returning from war (Bumiller February 26, 2009; Syrmopoulos January 31, 2013). While all of this information can technically be accessed by a persistent netizen, these frictions reduce citizens' awareness of the topics they intend to obscure, decreasing opposition to government and the policies of organized interests.

Despite the fact that information flooding and friction occur in similar ways in the U.S. as in China, it likely is somewhat more difficult for U.S. groups to create friction because organized groups have fewer restrictions, and therefore competition between groups in any given area is more widespread. As I have argued in this manuscript, in areas where much competition exists between groups, information is more easily leaked to the public. If this is the case, it implies that freedom of information is not as much driven by technology, but originates within the proliferation of civil society, interest group competition, and political competition.

6.2 Implications for Policy

This manuscript should heighten more than assuage concerns about information freedom in the digital age. Whereas much information is free in that it is possible for consumers to access it, my findings imply that free information is not about the *possibility* of access, but rather the *degree* of information accessibility. As long as organized groups and governments can prioritize information for consumers and as long as this prioritization is relatively invisible, consumer choice may not matter as much as the costs of information to consumers.

This should be even more worrying as governments move to make their censorship efforts more

invisible to consumers. China, in particular, has begun to make its censorship less noticeable, concealing censorship to bloggers, by allowing only authors of blogs to see censored posts that they have written, and making search filtering less apparent by removing error messages and simply refiltering search results. My manuscript implies that the observability of censorship alerts consumers to information manipulation, inspiring them to seek out more information. However with invisible censorship, most consumers will simply respond to the costs associated with accessing information. More investment needs to be made in understanding and communicating what topics are being censored in authoritarian regimes and the determinants of the accessibility of information to the public in autocracies and democracies. Only by creating awareness of manipulation can citizens respond to counteract its effects.

Bibliography

- “An insider’s account of the 50 Cent Party.” May 12, 2011b. *China Media Bulletin: Issue No. 22*, Freedom House. <http://www.freedomhouse.org/article/china-media-bulletin-issue-no-22/#2>.
- “Chasing the Chinese dream.” May 4, 2013. *The Economist*. <http://www.economist.com/news/briefing/21577063-chinas-new-leader-has-been-quick-consolidate-his-power-what-does-he-now-want-his>.
- “Chinese public want to see changes in monitoring pollution.” December 9, 2011a. *Taipei Times*. <http://www.taipeitimes.com/News/world/archives/2011/12/09/2003520325>.
- Ada, Sean, Henry Farrell, Marc Lync, John Sides and Deen Freelon. 2012. “Blogs and Bullets: New Media and Conflict after the Arab Spring.”. <http://j.mp/WviJPK>.
- Albayrak, Ayla and Joe Parkinson. September 16, 2013. “Turkey’s Government Forms 6,000-Member Social Media Team Volunteers to Promote Ruling Party’s Perspective in Sphere Dominated by Protesters.” *The Wall Street Journal*. <http://online.wsj.com/news/articles/SB10001424127887323527004579079151479634742>.
- Allison, Graham T and Morton H Halperin. 1972. “Bureaucratic politics: A paradigm and some policy implications.” *World Politics* 24(S1):40–79.
- Armstrong, Charles K. 2006. *Korean society: civil society, democracy and the state*. Routledge.
- Bamman, D., B. O’Connor and N. Smith. 2012. “Censorship and deletion practices in Chinese social media.” *First Monday* 17(3-5).
- Bellamy, Christine and Charles D Raab. 1999. “Wiring-up the deck-chairs?” *Parliamentary Affairs* 52(3):518–534.
- Bellin, Eva. 2012. “Reconsidering the Robustness of Authoritarianism in the Middle East: Lessons from the Arab Spring.” *Comparative Politics* 44(2):127–149.
- Besley, Timothy and Andrea Prat. 2006. “Handcuffs for the grabbing hand? The role of the media in political accountability.” *American Economic Review* 96(3):720–736.

BIBLIOGRAPHY

- Blanchard, Ben. July 8, 2012. "China pollution protest ends, but suspicion of government high." *Reuters*. <http://uk.reuters.com/article/2012/07/08/us-china-pollution-idUKBRE8670HP20120708>.
- Blei, David M, Andrew Y Ng and Michael I Jordan. 2003. "Latent dirichlet allocation." *The Journal of Machine Learning Research* 3:993–1022.
- Boyle, James. 1997. "Foucault in cyberspace: Surveillance, sovereignty, and hardwired censors." *University of Cincinnati Law Review* 66:177.
- Bradsher, Keith. July 4, 2012. "Bolder Protests Against Pollution Win Project's Defeat in China." *The New York Times*. <http://www.nytimes.com/2012/07/05/world/asia/chinese-officials-cancel-plant-project-amid-protests.html>.
- Brady, Anne-Marie. 2009. *Marketing dictatorship: Propaganda and thought work in contemporary China*. Rowman & Littlefield.
- Browning, Graeme. 2002. *Electronic democracy: Using the Internet to transform American politics*. Information Today, Inc.
- Bumiller, Elisabeth. February 26, 2009. "Pentagon to Allow Photos of Soldiers' Coffins." *New York Times*. <http://www.nytimes.com/2009/02/26/us/26web-coffins.html>.
- Cagé, Julia. 2013. "Media Competition, Information Provision and Political Participation."
- Chen, Wang, Wang Shanshan, Ren Zhongyuan and Zhu Yishi. February 19, 2013. "Dirty Business for China's Internet Scrubbers." *CNBC*. <http://www.cnbc.com/id/100472398>.
- Chen, Xi. 2012. *Social Protest and Contentious Authoritarianism in China*. New York: Cambridge University Press.
- Cohen, J Elizabeth and Joseph J Amon. 2011. "Lead poisoning in China." *The Lancet* 378(9803):e3.
- Coleman, Stephen and John Gotze. 2001. *Bowling together: Online public engagement in policy deliberation*. Hansard Society London.
- Dahlberg, Lincoln and Eugenia Siapera. 2007. *Radical democracy and the Internet: Interrogating theory and practice*. Palgrave Macmillan.
- Damm, Jens. 2007. "The Internet and the fragmentation of Chinese society." *Critical Asian Studies* 39(2):273–294.

BIBLIOGRAPHY

- Denyer, Simon. August 2, 2013. "In China, Communist Party takes unprecedented step: It is listening." *The Washington Post*. http://www.washingtonpost.com/world/in-china-government-mines-public-opinio/n/2013/08/02/33358026-f2b5-11e2-ae43-b31dc363c3bf_story.html.
- DeSilver, Drew. December 2, 2013. "China has more internet users than any other country." *Pew Research Center*. <http://www.pewresearch.org/fact-tank/2013/12/02/china-has-more-internet-users-than-any-other-country/>.
- Djankov, Simeon, Caralee McLiesh, Tatiana Nenova and Andrei Shleifer. 2001. Who owns the media? Technical report National Bureau of Economic Research.
- Edmonds, Rick, Emily Guskin, Tom Rosenstiel and Amy Mitchell. 2011. "Newspapers: by the numbers." *The State of the News Media 2011*.
- Efstathiou, Jim and Mark Drajem. June 6, 2013. "Drillers Silence Fracking Claims With Sealed Settlements." *Bloomberg*. <http://www.bloomberg.com/news/2013-06-06/drillers-silence-fracking-claims-with-sealed-settlements.html>.
- Etzioni-Halevy, Eva. 2013. *Bureaucracy and Democracy*. Vol. 7 Routledge.
- Ferdinand, Peter. 2000. "The Internet, democracy and democratization." *Democratization* 7(1):1–17.
- Galvis, Angela Fonseca, James M Snyder Jr and BK Song. 2012. "Newspaper market structure and behavior: Partisan coverage of political scandals in the US from 1870 to 1910."
- Gentzkow, Matthew and Jesse M Shapiro. 2008. "Competition and Truth in the Market for News." *The Journal of Economic Perspectives* pp. 133–154.
- Greve, Louisa. 2013. "The Troubled Periphery." *Journal of Democracy* 24(1):73–78.
- Grossman, Lawrence K. 1995. *The electronic republic: Reshaping democracy in the information age*. Viking Penguin.
- Gustin, Sam. March 25, 2010. "Google's Sergey Brin Leads Charge Against Chinese Web Censorship." *Daily Finance*. <http://www.dailyfinance.com/2010/03/24/googles-sergey-brin-leads-charge-against-chinese-web-censorship/>.
- Han, Enze. 2013. *Contestation and adaptation: the politics of national identity in China*. Oxford University Press.
- Harwit, Eric and Duncan Clark. 2001. "Shaping the internet in China. Evolution of political control over network infrastructure and content." *Asian Survey* 41(3):377–408.

BIBLIOGRAPHY

- Hassanpour, Navid. 2011. "Media disruption exacerbates revolutionary unrest: Evidence from Mubaraks natural experiment." *APSA 2011 Annual Meeting Paper* .
- He, Yinan. 2007. "History, Chinese nationalism and the emerging Sino-Japanese conflict." *Journal of Contemporary China* 16(50):1-24.
- Hille, Kathrin. October 19, 2010. "Google's China market share continues to fall." *Financial Times*. <http://www.ft.com/intl/cms/s/2/d14f6800-db87-11df-ae99-00144feabdc0.html#axzz2zzrByG70>.
- Hoelzle, Urs. January 2012. "The Google Gospel of Speed." *Google Think Insights*. <http://www.thinkwithgoogle.com/articles/the-google-gospel-of-speed-urs-hoelzle.html>.
- Huang, Haifeng and Zhi Li. 2013. "Propaganda and Signaling." *Available at SSRN 2325101* .
- Hughes, Christopher R and Gudrun Wacker. 2003. *China and the Internet: Politics of the digital leap forward*. Routledge.
- Jin, Qiu. 1999. *The culture of power: the Lin Biao incident in the Cultural Revolution*. Stanford University Press.
- Kalathil, Shanthi and Taylor C Boas. 2010. *Open networks, closed regimes: The impact of the Internet on authoritarian rule*. Carnegie Endowment.
- Keohane, Robert O and Joseph S Nye Jr. 1998. "Power and interdependence in the information age." *Foreign Affairs* pp. 81-94.
- King, Gary, Jennifer Pan and Margaret E. Roberts. 2013a. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* 107:1-18. <http://j.mp/LdVXqN>.
- King, Gary, Jennifer Pan and Margaret Roberts. 2013b. "A Randomized Experimental Study of Censorship in China." Copy at <http://j.mp/16Nvzge>.
- LaFraniere, Sharon. June 15, 2011. "Lead Poisoning in China: The Hidden Scourge." *The New York Times*. <http://www.nytimes.com/2011/06/15/world/asia/15lead.html?pagewanted=all>.
- Latham, Kevin. 2007. "SMS, communication, and citizenship in China's information society." *Critical Asian Studies* 39(2):295-314.
- Lessig, Lawrence. 1999. *Code: And other laws of cyberspace*. Basic Books (AZ).
- Lieberthal, Kenneth. 1995. *Governing China: from revolution through reform*. WW Norton New York.

BIBLIOGRAPHY

- Lin, Lilian. July 30, 2012. "Qidong Protest Prompts Anti-Japan Sentiment." *The Wall Street Journal*. <http://blogs.wsj.com/chinarealtime/2012/07/30/qidong-protest-prompts-anti-japan-sentiment/>.
- Liphshiz, Cnaan. January 19, 2009. "Israel recruits 'army of bloggers' to combat anti-Zionist Web sites." *Haaretz*. <http://www.haaretz.com/print-edition/news/israel-recruits-army-of-bloggers-to-combat-anti-zionist-web-sites-1.268393>.
- Lorentzen, Peter. 2013. "China's Strategic Censorship." *American Journal of Political Science*.
- Lorentzen, Peter L, Pierre F Landry and John K Yasuda. 2010. "Transparent authoritarianism." *Analysis of Political and Economic Barriers to Greater Government Transparency in China*.
- Lotan, Gilad, Erhardt Graeff, Mike Ananny, Devin Gaffney, Ian Pearce and Danah Boyd. 2011. "The revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions." *International Journal of Communication* 5:1375–1405.
- Lynch, Daniel C. 1999. *After the propaganda state: Media, politics, and "thought work" in reformed China*. Stanford University Press.
- MacKinnon, Rebecca. 2011. "Chinas networked authoritarianism." *Journal of Democracy* 22(2).
- McCombs, Maxwell E and Donald L Shaw. 1972. "The agenda-setting function of mass media." *Public opinion quarterly* 36(2):176–187.
- Milward, Stephen. June 30, 2011. "Google+ Not Actually Blocked in China, Just Being Slowly Throttled." *TechAsia*. <http://www.techinasia.com/google-plus-china/>.
- Mullainathan, Sendhil and Andrei Shleifer. 2005. "The market for news." *American Economic Review* pp. 1031–1053.
- Nathan, Andrew J. 1973. "Factionalism Model for CCP Politics, A." *China Quarterly* p. 34.
- National Academy for Propaganda Cadres, ed. 2011. *Xuanchuan sixiang wenhua gongzuo: anli xuanbian [Propaganda, thought and cultural work: selected cases]*. Beijing: Study Press.
- Nichols, John and Robert W McChesney. 2009. "The death and life of great American newspapers." *The Nation* 18:2009.
- Olson, Mancur. 2009. *The logic of collective action: public goods and the theory of groups*. Harvard University Press.
- Perry, Elizabeth J. 2013. "Cultural Governance in Contemporary China: Re-orienting Party Propaganda."

BIBLIOGRAPHY

- Potters, Jan and Frans Van Winden. 1992. "Lobbying and asymmetric information." *Public choice* 74(3):269–292.
- Qin, Bei, Yanhui Wu and David Strömberg. 2012. "The Determinants of Media Bias in China." Working Paper.
- Ratkiewicz, J., F. Menczer, S. Fortunato, A. Flammini and A. Vespignani. 2010. Traffic in social media II: Modeling bursty popularity. In *Social Computing, 2010 IEEE Second International Conference*. IEEE pp. 393–400.
- Regalbuto, Eric. 2012. "Inferno in the Land of Snows: A Holistic Investigation of Tibetan Self-Immolation Through a Tibetan Perspective."
- Roberts, Margaret E, Brandon M Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Gadarian, Bethany Albertson and David Rand. Forthcoming. "Structural topic models for open ended survey responses." *American Journal of Political Science*.
- Roberts, Margaret E, Brandon M Stewart and Edoardo M Airolidi. 2013. Structural Topic Models. Technical report Working paper.
- Shen, Xiao-ming, John F Rosen, Di Guo and Sheng-mei Wu. 1996. "Childhood lead poisoning in China." *Science of the Total Environment* 181(2):101–109.
- Shih, Victor. 2008a. "Nauseating Displays of Loyalty: Monitoring the Factional Bargain through Ideological Campaigns in China." *The Journal of Politics* 70(04):1177–1192.
- Shih, Victor C. 2008b. *Factions and finance in China*. Cambridge: Cambridge University.
- Shirk, Susan L. 2011. *Changing Media, Changing China*. New York: Oxford University Press.
- Siebert, Fred Seaton. 1956. *Four theories of the press: The authoritarian, libertarian, social responsibility, and Soviet communist concepts of what the press should be and do*. University of Illinois Press.
- Sisci, Francesco. November 1, 2012. "Wen versus Bo gets more explosive." *Asia Times*. <http://www.atimes.com/atimes/China/NK01Ad01.html>.
- Slegg, Jennifer. July 9, 2013. "Google's Search Market Share in China Falls to Just 3%." *Search Engine Watch*. <http://searchenginewatch.com/article/2280420/Googles-Search-Market-Share-in-China-Falls-to-Just-3>.
- Stevenson, Christopher. 2007. "Breaching the Great Firewall: China's Internet Censorship and the Quest for Freedom of Expression in a Connected World." *BC Int'l & Comp. L. Rev.* 30:531.

BIBLIOGRAPHY

- Stockmann, Daniela. 2012. *Media commercialization and authoritarian rule in China*. Cambridge University Press.
- Syrmopoulos, Jay. January 31, 2013. "Media Blacks Out New Snowden Interview The Government Doesn't Want You to See." *Ben Swann* <http://benswann.com/media-blacks-out-new-snowden-interview-the-government-doesnt-want-you-to-see/>.
- Tang, Kevin. September 30, 2013. "For Chinese Bloggers, Going Viral Can Mean Jail Time." *Buzzfeed*. <http://www.buzzfeed.com/kevintang/culling-the-big-vs-china-cracks-down-on-celebrity-bloggers>.
- Taubman, Geoffrey. 1998. "A not-so world wide web: the Internet, China, and the challenges to nondemocratic rule." *Political Communication* 15(2):255–272.
- Tolbert, Caroline J and Ramona S McNeal. 2003. "Unraveling the effects of the Internet on political participation?" *Political Research Quarterly* 56(2):175–185.
- Tong, Jingrong. 2011. *Investigative journalism in China: journalism, power, and society*. Bloomsbury Academic.
- Tong, Jingrong and Colin Sparks. 2009. "Investigative journalism in China today." *Journalism Studies* 10(3):337–352.
- Trippi, Joe. 2005. *The revolution will not be televised: Democracy, the Internet, and the overthrow of everything*. Regan Books.
- Watt, Louise. December 8, 2011. "China's pollution data shrouded in official fog." *Business Week*. <http://www.businessweek.com/ap/financialnews/D9RGF0EO1.htm>.
- Wedeen, Lisa. 1999. *Ambiguities of domination: Politics, rhetoric, and symbols in contemporary Syria*. University of Chicago Press.
- Weiss, Jessica Chen. 2013. "Authoritarian Signaling, Mass Audiences, and Nationalist Protest in China." *International Organization* 67(1).
- Weiss, Jessica Chen. 2014. *Powerful Patriots: Nationalist Protest in Chinas Foreign Relations*. Oxford University Press.
- Wong, Edward. April 1, 2013. "Air Pollution Linked to 1.2 Million Premature Deaths in China." *New York Times*. http://www.nytimes.com/2013/04/02/world/asia/air-pollution-linked-to-1-2-million-deaths-in-china.html?_r=1&.
- Yang, Guobin. 2009. *The Power of the Internet in China: Citizen Activism Online*. New York: Columbia University Press.

BIBLIOGRAPHY

- Yang, Guobin. 2014. "Internet Activism & the Party-State in China." *Daedalus* 143(2):110–123.
- Yu, Frederick Teh-Chi. 1964. *Mass Persuasion in Communist China*. London & Dunmow; printed in USA.
- Zhao, Yuezhi. 2008. *Communication in China: Political economy, power, and conflict*. Rowman & Littlefield Publishers.
- Zheng, Yongnian. 2007a. *Technological Empowerment: The Internet, State, and Society in China*. Stanford, CA, USA: Stanford University Press.
- Zheng, Yongnian. 2007b. *Technological empowerment: The Internet, state, and society in China*. Stanford University Press.
- Zhu, Tao, David Phipps, Adam Pridgen, Jedidiah R Crandall and Dan S Wallach. 2013. "The Velocity of Censorship: High-Fidelity Detection of Microblog Post Deletions." *arXiv preprint arXiv:1303.0597*.
- Zhu, Weijing. April 21, 2014. "Chinese internet censors bribed to delete posts." *The World of Chinese*. <http://www.theworldofchinese.com/2014/04/chinese-internet-censors-bribed-to-delete-posts-for-local-governments/>.