Fear or Friction? How Censorship Slows the Spread of Information in the Digital Age *

Margaret E. Roberts†

August 13, 2014

Abstract

What techniques do governments use to prevent the spread of information, and when are they effective? I examine two alternative mechanisms through which censorship could slow the spread of online information. On one hand, censorship could create awareness of sensitive topics, signaling government intent and inducing “fear” so that people self-censor on topics the government deems objectionable. Alternatively, censorship can create “information friction”, making information slightly more difficult and costly to obtain, but keeping the act of censorship itself as invisible as possible. I examine whether each mechanism is effective in slowing the spread of information using a novel experimental design which creates awareness of censorship among consumers of social media in China, in addition to a unique observational dataset that explores how bloggers change topics in reaction to censorship and measures the spread of information about protest events in China. I find that awareness of censorship on the part of consumers and producers of blogposts does not deter the spread of information and instead often undermines government legitimacy and induces information seeking. However, small increases in the cost of information that are not apparent to the public are influential at stopping the spread of information to large, politically meaningful groups of people. My findings help clarify how citizens are affected by censorship, explain some recent shifts in the Chinese government’s censorship policies, and have broader implications for authoritarian resilience and freedom of information in the digital age.

*My thanks to Joan Cho, Jeff Frieden, Shelby Grossman, Andrew Hall, Gary King, Jennifer Pan, Elizabeth Perry, Amanda Pinkston, David Steinberg, Brandon Stewart, Chiara Superti, Dustin Tingley, Jessica Weiss and Vanessa Williamson and the participants in Harvard’s Comparative Politics Workshop and UC Merced’s Understanding Politics Speaker Series for helpful comments.
†Assistant Professor, Department of Political Science, USCD; http://scholar.harvard.edu/mroberts/home
Introduction

Government censorship and control over the Internet is well-documented. Out of 60 countries surveyed, Freedom House estimates that only 17% of countries have no political censorship, and over one-third of the countries have intimidated bloggers, blocked or filtered websites, or launched cyberattacks against critics of the regime (Kelly et al., 2012). While many originally thought the Internet would make censorship defunct (Taubman, 1998; Yang, 2009) these controls are by no means getting more lenient – according to Freedom House, over half of the countries surveyed had lower Internet freedom in 2012 than the year before.

Even though many countries around the world pay billions of dollars to censor the web, there is still much scholarly debate over whether governments can influence the spread of information in the age of the Internet (Lessig, 1999; Zheng, 2007; Hughes and Wacker, 2003). This debate speaks to larger questions about the Internet’s role in the development and success of social movements and whether states can maintain power in a digital age of inter-connectedness. However, data to measure both state censorship and the spread of information to the public is difficult to obtain, and therefore whether censorship can stop the spread of digital information is difficult to test empirically. While a few scholars have studied the influence of complete Internet blackouts (Kern and Hainmueller, 2009; Hassanpour, 2011), to my knowledge none so far have looked at the influence of more common, surgical censorship methods such as government removal of individual posts from the Internet.

In this paper, I use experiments in Beijing and collect data on posts and censorship of social media in China to show that the primary mechanism through which censorship is effective in slowing the spread of information is by creating information friction, or by making access to information less convenient. I find, unexpectedly, that fear produced by awareness of government censorship does not slow the spread of information, in other words, anxiety caused by observation of censorship does not cause self-censorship among average citizens in China. In fact, in some cases, awareness of government censorship instead undermines the legitimacy of the government censorship laws.
In the sections below, I begin by defining *information friction* and outlining two potential mechanisms through which censorship could affect the flow of information. I then describe the experiments and data I use to test these theories. After explaining the results of each of these tests, I conclude.

**Mechanisms Through Which Censorship Could be Effective**

I extend the work of others on the influence of censorship (Lessig, 1999; Zheng, 2007; Hughes and Wacker, 2003) by establishing a theoretical foundation for the ways in which censorship could be effective in stopping the spread of information on the Internet. Instead of focusing on how the Internet affects the overall environment in which people interact, I develop a theory based on the individual by trying to understand the logic behind individual interaction with information on the Internet.

In doing so, I hypothesize that there are two different mechanisms through which censorship could be effective in stalling the spread of information at the individual level. What distinguishes these two mechanisms is whether the netizen is aware that the information they are interacting with is being censored by the government. The first mechanism is *fear*, where the observation of censorship signals sensitive topics to bloggers. In response, netizens might decide not to read or write about a particular topic because they fear government reprimands if they do so. This is a direct mechanism through which censorship could function – because people observe censorship, they decide not to inform themselves or inform others about a particular issue, thus stalling the spread of information.

Second is *information friction*, or cost of access, which does not require an awareness of censorship. The selective removal of blogposts online, website blocking, or denial of service attacks may raise the costs of access sufficiently so that consumers of information either do not happen upon that information or become frustrated accessing it. These netizens may be completely unaware that censorship is occurring, but their behavior is still affected by censorship.

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 induce fear, regardless of technology (Keohane and Nye Jr, 1998). Boyle (1997) speculates that state control of the Internet will come through fear of unknown and sporadic controls, which will induce self-censorship. Hughes and Wacker (2003) argue that the perception of surveillance in Chinese society and a few high-profile arrests are more important than actual censorship in stopping the spread of information.

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 a clear 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 jailings 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 even detained after he claimed the police in Gansu province had been complicit in a death at a karaoke bar.\(^1\)

Even though censorship can easily produce fear, it is obviously government-driven, and is a signal of government efforts at the suppression of information. When a person sees that her post or another post has been censored, she knows that the government objects to its content, and that there is a certain (unknown) probability that continuing to write on that topic will make her life more difficult. However, if the person opposes censorship, this anxious reaction might be overpowered by dissatisfaction with government censorship policies, and might instead induce more negative views of the government, inspiring her to continue writing about the sensitive topic in order to undermine censorship.

Censorship could also signal government weakness. Censorship is an indication that the government has something to hide, and therefore may signal to the observer of censorship that the government is weaker than they previously believed. Indeed, political sci-

\(^1\)http://www.buzzfeed.com/kevintang/culling-the-big-vs-china-cracks-down-on-celebrity
entists have indicated that censorship blackouts may undermine government efforts, for example Hassanpour (2011) presents evidence that complete Internet blackout in Egypt during the Arab Spring undermined government legitimacy because of its observability. Indications of weakness could embolden writers to continue to post, despite the dangers of doing so (Huang and Li, 2013).

Information friction, in contrast, does not require awareness of government censorship. Information friction is created when censorship makes information more difficult to access. Even small costs of access could potentially have large implications. Market researchers have found that small delays online can have huge effects on consumers of information. Research has shown that even fractions of a second on website loading time can drive impatient consumers away (Cheung and Lee, 2005). Google market research shows that slowing search by 400 milliseconds creates a .44 percent drop in search volume.\(^2\) Other research shows that slow internet is aggravating enough that 80\% of internet users will get frustrated and leave Google’s Youtube if loading a video stalls. Search rankings are well-known to have huge effects on the number of hits on a website. These findings likely translate from the commercial to the political world. Even if censors cannot scrub information completely from the web, inducing small costs to access may mean that many fewer people come across the information, whether they are aware it is being censored or not.

This evidence aside, information friction has been written off by many scholars of the Internet who argue that information travels too quickly in the age of information to be slowed down solely by bumbling government censors (Garrett, 2006; Yang, 2009; Shambaugh, 2007; Benkler, 2006). Censoring the Internet, according to former President Bill Clinton, is like trying to “nail jello to the wall”. Either information is kept completely secret, in which case censorship is effective, or it is quickly leaked and then easily known to everyone.

\(^2\)http://www.thinkwithgoogle.com/articles/the-google-gospel-of-speed-urs-hoelzle.html
Outline of Results

The mechanisms of fear and friction are such that both or neither could be important to the spread of information. If Internet censorship is not effective at all, then neither friction nor the awareness of censorship stalls the spread of information. However, if censorship is effective, then both mechanisms could be at play in slowing the spread of information.

In this article, I study typical consumers and producers of blogs in China and find that censorship does have some effectiveness. However, the efficacy of censorship does not play out through fear for these typical social media users, but rather through small increases in the cost of access. I show that when netizens observe censorship, it piques their interest in the censored topic, causing them to seek out more information, instead of less, on that topic. Bloggers and other social media producers do not shift topics in response to censorship. They do not slow their production of blogposts after censorship, and they do not become more supportive of the state after censorship.

On the other hand, slight delays in censorship induce large changes in the spread of information about similar events. Using 120 self-immolation events in Tibet, I show that the difference in the spread of information between events is related to the schedule of the censors. Because these events are uniformly censored in China, self-censorship should be constant across them, implying that information friction, not the knowledge of which topics are censored is the primary way through which censorship is effective.

Data

I use three distinct datasets to understand the efficacy of censorship and the mechanism through which censorship operates. In each test, my aim is to study typical, rather than high-profile producers and consumers of blogposts in China.\(^3\) 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 and sentiment with which bloggers write. In the second test, I hold information friction constant and randomize fear in an experiment with consumers.

\(^3\)High-profile bloggers have received more attention in the news media, but do not create the majority of the content on the Chinese blogosphere.
of blogposts in China. I study how the observation of censorship influences participants subsequent consumption of blogposts. In the third test, I study 120 self-immolation events in Tibet, which all have high censorship rates. I examine how the day within the week of the event influences the spread of information about the event. I find that on the weekends, when censorship is more delayed, information spreads more quickly, implying that small costs of access to information have important implications for the spread of information online.

Limitations

As I discuss more below, while these three tests all seek to understand different aspects of how fear and friction influence the spread of information, they have some important 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, the lab is always somewhat artificial. 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 fear and friction.

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 come across censorship only in cases where links are broken or forums have missing threads, producers of blogposts are likely to come across censorship if or when they are censored. Bloggers will likely notice that their blogpost has gone missing, particularly because censorship typically occurs quickly, usually within one day (King, Pan and Roberts, 2013), and therefore their most recent
posts are likely to be the posts that have been removed.\(^4\)

Bloggers in China have a history of being arrested or intimidated by the government as a result of the content that they produce and therefore are likely to suffer higher costs for their actions than consumers of blogposts. Famous activist and artist Ai Weiwei was arrested in 2011 ostensibly on tax evasion, but likely for 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 and new rules restricting “rumors” and other types of political discussion have been implemented that give the government more power to arrest bloggers. These highly publicized arrests were based on a new law where if a rumor or other illegal post was re-shared or viewed enough times, the blogger could be jailed for up the three years. This law puts anyone at risk of being jailed, including the middle-schooler discussed above.

Overall, we would expect that bloggers, more than consumers of blogposts, would be deterred from writing about sensitive political issues after experiencing censorship. Censorship signals the type of information the government finds objectionable, and therefore might make bloggers inclined to stop talking about that topic or shift their topic to a less objectionable one. They might also change their online commentary about the government, if they are afraid of being arrested.

**Data and Methods for Testing Blogger Self-Censorship**

To study self-censorship of bloggers, I sampled 2,000 bloggers whose blogs existed on a particular Internet content provider in China that appeared in King, Pan and Roberts (2013) dataset. I chose this sample because I had collected data on all of the blogposts these bloggers has written since 2010, before censorship. These bloggers also represented a wide range of blogger-types, some with very few followers, some with very many followers, and of all ages and genders. I then returned to each of these 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

\(^4\)In increasing numbers of cases, Internet content providers try to hide censorship from bloggers, by allowing them to see 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.
bloggers had been “treated” with censorship.

To study how censorship changes behavior, I focus on estimating the within blogger change in topic after censorship. That is, in the period right before the blogger experienced censorship, does the blogger talk about different topics than the period right 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 between bloggers is quite high, and therefore a comparison between bloggers who have been censored and those who were never censored would reflect what topics were censored, not how bloggers are affected by censorship. Although what types of topics are censored is an interesting question in itself, I reference the reader to King, Pan and Roberts (2013) for discussion of this question and instead here focus on how individuals are affected by censorship.

I also removed all bloggers whose blogs were now completely missing, as I could not tell if they were missing due to government intervention or the blogger’s own choice. After sub-setting, I was left with 516 bloggers of interest. From each of the 516 bloggers, I randomly selected a censored post. I then took the five blogposts directly preceding the censored posts, and the five blogposts directly following the censored post.\(^5\) In the end, my dataset included with 5,199 blogposts, from 516 bloggers.

**The Structural Topic Model**

To analyze the blogposts, I used a new method for text analysis called the Structural Topic Model (Roberts et al., 2013; Roberts, Stewart and Airoldi, 2013). 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.

\(^5\)If the censored post had fewer than five posts preceding or following it, I took the maximum number of posts I could get before and following.
STM is also an appropriate choice for this application because it allows for the inclusion of blogger-specific fixed effects and therefore allows within-blogger estimation of how bloggers change topics in reaction to censorship. I include blogger-specific fixed effects as well as an indicator for whether the blogpost was written before or after censorship in the STM model.\textsuperscript{6}

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 and therefore 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 censorship does not deter bloggers from writing, and that any anxiety produced by censorship is perhaps offset by countervailing considerations or emotions.

Results

I estimate a 50 topic model on 5,199 blogposts. The main dimension along which topics orient themselves is between two broad categories: topics relating to politics, and topics relating to personal life events. An overview of the topics and their prevalence within the dataset can be found in Figure 1. Political topics include everything from party politics to international relations. Personal topics range from diary-like topics explaining what the person did today, to writing about love, family, and work. Other types of topics range from computers (how to download something, what operating system to use) to

\textsuperscript{6}I 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.
art (celebrities, poetry, etc) to political history (from Ancient Chinese history to Mao) to economics (exchange rates, small business, and the stock market). An individual blogger is likely to talk about several different topics throughout the time-period they are blogging.

Figure 2 shows a correlation plot of the topics of interest. 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 colored by larger topical themes, including government, economics, history, art, and personal topics. As you can see, 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, indicating that they are the two main dimensions of conversation on the blogosphere.

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 over 70% about a political topic. This removed bloggers whose censored post could have been due to spam, or personal topics, which might be more likely to have been removed by the blogger themselves.

Overall, bloggers whose censored post was about a political topic are much more likely to talk about economic and political topics in their other posts than those whose censored post was not mainly about an economic or political topic. Figure 3 shows the most common topics for this subset of bloggers. Within these blogs writing, I also looked in detail at the censored posts. The most common censored blogposts in this sample are posts related to the downfall of Bo Xilai and anti-Japanese protests. This is consistent with King, Pan and Roberts (2013), who find that the Chinese government censors blogposts about collective action, since these two events are both collective action events.

I use this subset of bloggers to determine whether bloggers have different topic proportions before and after censorship. Bloggers do not exhibit a significant shifts on any individual topic. Figures 4 and 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.

---

7I removed topics that were spam or contained only general words
Figure 1: Top Topics over all of the blogposts within the study.
Only when I aggregate the political and life topics, do I find that bloggers exhibit a 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 out of topics related to politics. However, this shift is quite small – bloggers decrease the political proportions of their post on average 1% and increase the personal life proportion of their post on average .1%. These are extremely small effects, that are substantively insignificant. While fear of government backlash might influence the writing of one in 100 bloggers, it does not seem to have a substantively meaningful influence. Censorship also does not 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
Figure 3: Topic ordering for bloggers whose censored post was political.
Figure 4: Shift in Political Topics, Before and After Censorship
Figure 5: Shift in Life Topics, Before and After Censorship
Figure 6: Relationship between post-censorship posts and sentiment.

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 propensity for further censorship, which one might expect if bloggers shifted to topics that were less objectionable to the government. On average blog posts after have 3% less censorship 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 6, I found no significant difference between critical, neutral and supportive categories before versus after censorship. Censorship, it seems, does not change the sentiment of the blogger toward the government.
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 matters in 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 as above: bloggers did not change the topics they discussed before and after censorship.

Implications for the Spread of Information

The finding that bloggers do not change topics or alter their commentary of the government in reaction to the awareness of censorship implies that fear produced by censorship does not contribute to slowing the spread of information. 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. Awareness of censorship, for the average blogger, does not slow the spread of information.

Test Two: An Experimental Study of Consumers of Blogposts

While censorship in China is less obvious 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. When the consumer clicks the link, the user will be taken to an error page instead of the page with the original content. An example error page is displayed in Figure 7.

Another way consumers of social media can observe censorship is within a forum or a discussion thread. When posts are censored within a forum, the content of the individual comment is removed, but the rest of the conversation still exists. Therefore, the user will
see a removed post where the original post once stood. This can also occur on Sina Weibo (the Chinese equivalent of Twitter), where sometimes when an individual post is removed, the rest of the thread stays up.

When consumers come across a censored page, they often guess as to the topic of the censored post because of the context of the censorship. For example, if the user clicked on a link to a censored page, the title of the link will often give an indication of what the original blogpost was about. Within a forum, consumers of information will see the discussion surrounding the missing post, and therefore may be able to guess the post’s content.

Consumers of blogposts in China are often suspicious that their information is being tracked online and that they are subject to surveillance. Many Internet content providers now require government IDs or a phone number for users to sign up for an online ID. 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 thwarted by Internet users, shortly after local governments began mandating similar software under different names (MacKinnon, 2011).

In an age where online surveillance is common, if consumers of information sometimes happen upon censored blogposts or other types of social media, how do they react? If the consumer is primarily acting on the fear produces by the censored post, interaction with a censored blogpost will cause the reader to avoid further interaction with the topic.
On the other hand, if censorship creates countervailing emotions, such as indignation or curiosity, it might instead pique the interest of the reader, interesting them in reading more about the topic and undermining the legitimacy of the government. 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.

**Design**

I conducted an experiment on self-censorship for 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 were given access to the Internet. They were provided with a list of blogposts in a blog-aggregator. They were told 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 about on-going events in China: protests over the construction of para-exlyne (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 within China, more sensitive topics could not be used, but pre-testing of the experiment suggested that these were all topics Chinese citizens expect to be sensitive and expected could be censored. 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 about.

For the treated group, a randomly selected blogpost 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.

The question of interest was – how does censorship of that blogpost influence the topics the reader selects next? Is the reader less likely to pick a blogpost on that topic because they have 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, however, we would expect participants to be equally or even more likely to click on the censored topic.

**Results**

The treatment had its intended influence on the subjects – when asked what the error page indicated, most (two-thirds) of the respondents said that the page was 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 PX protests were significantly more likely to say that the government censored information about protests than those not treated with
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 click on a post within the same topic next than those who came across uncensored posts, keeping the topic of the blogpost constant. Further, readers who come across censored posts were likely to click on more than one more new post about that topic, whereas readers who come across uncensored posts one average clicked on less censored posts related to protests.\footnote{Respondents 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 8: Treatment Effect of Censorship on the Probability on Clicking on a Post of the Same Topic Right After Treatment

Figure 9: Treatment Effect of Censorship on the Proportion of Posts Clicked on About the Same Topic After Treatment
than one new post (Figure 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 fact that the blogger had taken the page down.

More indicative, treated subjects had different views than control subjects of how much the Internet should be censored. 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 indicated that they thought censorship should be significantly lower than control subjects. This difference in opinion points to the fact that the observation of censorship is probably creating countervailing emotions that undermine the legitimacy of the government’s information laws.

The implication of these results is 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 interested readers in learning more about the topic, and created disillusionment with government censorship laws. Inducing awareness of censorship in this case, backfired to undermine government policy.

If readers are not deterred, but rather more interested by the awareness of censorship, observable censorship likely increases the spread of information rather than stalling it. Consumers are more likely to read about the topic, and might therefore be more likely to pass it on to their friends. They might be angered by censorship and emboldened to oppose government implementations of it. The evidence suggests that the mechanism has the opposite effect than expected.

**Test Three: Weekend vs. Weekday Self-Immolations in Tibet**

The previous two tests show that fear does not deter bloggers from writing or reading about sensitive topics. The evidence suggests that self-censorship cannot be the primary way through which censorship is effective in China. In this next test, I compare the spread of information across self-immolation events in Tibet from 2011-2013. Since such protests are uniformly censored, all considered very sensitive in China, and comparable in content, fear of government reprisal for reading and writing about such events should
be constant over the events. Holding fear constant, I examine how small changes in cost of access influence the spread of information about these protest events.

**Background of Self-Immolation Protests**

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 which the Chinese government opposes (Regalbuto, 2012).

These protest events represent a huge political 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 ethnic affairs and fighting independence movements are central to Chinese national security policy, the immolations are a direct challenge to the Chinese government. Self-immolations in other countries also have a history of causing 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).

To discourage such protest events, the government quickly responds to self-immolation events, removing the self-immolator as quickly as possible and increasing police presence to prevent protests in the aftermath. Police often punish villagers and families of Tibetans of 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).

News of Tibetan self-immolations are uniformly and quickly censored on social media websites. Unlike most political discussions, where one thread will generate thousands of re-shares in China, threads spreading news of Tibetan self-immolators are quickly cut off. Since information about self-immolations is scarce, social media accounts of self-immolations are often uncertain, for example, “I heard there was a self-immolation today
in Tibet. Is it true?”. News about specific self-immolation events are an example of a topic that is uniformly censored by all social media websites in China, and therefore fear of writing and reading about self-immolation events should be constant across events.

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

To study how censorship influences the spread of news surrounding self-immolation events, I collected a 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. Because my goal is to understand why certain self-immolations are discussed more than others, the sample was random and representative across time, meaning that spikes in social media postings during the time period are represented within the sample. I then checked to see which of these posts had been removed from the Internet.

Discussion about 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 am measuring how many posts were written about the event. However, depending on how quickly censorship occurs, many of these posts were only available to readers for a short period of time. The government censors many of the postings about self-immolations and are 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 receive more attention from bloggers than others – some have longer bursts, or more discussion about the event. Figure 10 shows the variation in bursts across the 120 self-immolation events between 2011 and 2013. Some events receive barely any attention at all, while others have a large amount of social media discussion associated with the event.

Why would some immolations receive more attention than others? It could be that the nature of the event was such that certain events received more attention from the public than others. The age of the immolator is usually something people note when discussing
Figure 10: Length of Bursts for 120 Self-Immolation Events between 2011 and 2013
self-immolation events, with younger immolators often being discussed with more grief than older immolators. Monks who self-immolate might have a larger network of followers, 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 censorship, 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 smaller the information available associated with 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 (2013), which was 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 (2013) 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 drastic, 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 likely to self-immolate on a weekend, and also are 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 11, showing a positive effect of an event happening on the weekend on the expected number of posts within the burst, and the model itself is included within the Appendix.

To verify that this effect was not a result of people simply writing more posts on the weekend, I downloaded a random sample of blogs from the same source mentioning Tibet during the same time period. These posts are largely not sensitive; most talk about economic opportunities in Tibet, traveling in Tibet, or Tibetan culture. Surprisingly, people are significantly less likely to write about Tibet on the weekend than during the weekday. This is even further evidence that the length of the burst is not due to general interest, but instead to differences in information control.

The fact that the speed of content filtering influences the number of posts about a self-immolation event indicates that small costs of access to information, such as the timing
Figure 11: Weekend Self-Immolations Have Longer Bursts Than Weekday Self-Immolations, Negative Binomial Model

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 influenced the number of people that know about an event. This suggests that information friction indeed slows down the spread of online information in China.
Discussion

Why would information friction be more effective than fear in preventing the spread of information? Whereas fear has a negative side-effect of damaging the reputation of the government and possibly causing a negative reaction, friction discourages netizens from reading a blogpost, and they are unaware of the topics of blogposts that they have not read. When netizens can not find information about a certain topic of interest, they often have no way of telling whether information is difficult to find because no one is talking about the event, or if information is difficult to find because the government is censoring it. Even though some netizens might run into the information before the government censors, friction prevents large groups of people from becoming aware of the event, therefore preventing collective action events and news from spreading.

This finding that information friction rather than fear is more effective in stopping the spread of information might explain recent changes in Chinese censorship policies. For example, one commonly used censorship method is called “search filtering”, where searching for a sensitive term within social media sites and search engines returns an error. Recently, many social media sites have made search filtering less apparent to the user by changed search filtering so that it returns a list of unobjectionable websites, instead of an error indicating that the search term is objectionable.

This could also explain why the Chinese government focuses its intimidation efforts on high-profile bloggers, rather than normal netizens. While this study studies typical netizens to understand the spread of the majority of information, it could be that high-profile bloggers exhibit high levels of self-censorship, since interviews with government officials suggest that bloggers with large numbers of followers are frequently visited by government police who intimidate them into self-censoring on sensitive topics. However, if even high profile bloggers react to this fear, this is paradoxically still consistent with a government strategy of friction. Whereas friction may be the main mechanism in stopping the transmission of news between regular citizens, since high-profile bloggers have large number of followers, intimidation of these writers has downstream friction effects on the spread of information than typical bloggers. This may explain why the government
focuses most of its intimidation efforts on those with high numbers of followers.

The finding also might explain why countries with less state capacity often revert to intimidation, while wealthier countries like China use technology that makes censorship less visible to the consumer. As tools and money to make censorship invisible increases, we should expect that governments will decrease move toward less observable forms of censorship. However, this has negative implications for netizens, as it will be more difficult for consumers of information to quickly identify government interference in information flow.

**Conclusion**

This paper tests how fear and information friction influence the spread of information online in China. It uses three unique datasets to do this: a randomized experiment to understand how fear influences consumers of blogposts in China, an observational dataset to understand how bloggers respond to censorship, and blogposts from 120 self-immolation events to understand how delays in censorship influence the spread of information. I find that fear has very little effect on consumers and producers’ interactions with blogposts, but small delays in censorship have large effects on the spread of information online. This suggests that the most pernicious forms of censorship are invisible to the consumer and that small, government-imposed delays in online information can be effective in stalling information flow.
Appendix

Table 1: Model of Correlation Between Weekend Self-Immolations and Length of Social Media Burst

<table>
<thead>
<tr>
<th></th>
<th>Outcome: Burst Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>(0.49)</td>
</tr>
<tr>
<td>Weekend</td>
<td>0.76*</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
</tr>
<tr>
<td>Age</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
</tr>
<tr>
<td>Monk</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
</tr>
<tr>
<td>Death</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>(0.40)</td>
</tr>
<tr>
<td>$\theta$</td>
<td>1.32*</td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
</tr>
<tr>
<td>N</td>
<td>111</td>
</tr>
</tbody>
</table>

Standard errors in parentheses
* indicates significance at $p < 0.05$
References


Huang, Haifeng and Zhi Li. 2013. “Propaganda and Signaling.” *Available at SSRN 2325101*.


