# Normalization of Censorship: Evidence from China*

Tony Zirui Yang†

June 23, 2021

**Abstract**

Previous research claims that public awareness of censorship will lead to backlash against the regime. However, surveys consistently find that Chinese citizens are apathetic toward or even supportive of government censorship. To explain this puzzle, I argue that citizens are subject to a process of normalization. Specifically, individuals become desensitized to censorship when the range of censored content expands beyond politically threatening topics like government criticism and collective action to other seemingly harmless non-political issues. Using a dataset of 15,872 censored articles on WeChat and an original survey experiment in China, I show that (1) a majority of censored articles are unrelated to politically threatening topics, and (2) respondents exposed to the censorship of both political and non-political content display less backlash toward the regime and its censorship policy than those who were only exposed to political censorship. My findings highlight normalization as a powerful tool of authoritarian control.

**Keywords:** Censorship, China, Normalization, Desensitization, Backlash, Authoritarian Control

Word Count: 10,157

*I thank Margit Tavits for her advice and generous support throughout this project. I also thank Afiq bin Oslan, Taylor Carlson, Haohan Chen, Taylor Damann, Haifeng Huang, Dahjin Kim, Christopher Lucas, Jordan H. McAllister, Bryant Moy, Benjamin S. Noble, Jennifer Pan, Juan J. Qian, Alexandra Siegel, and Jeremy Siow for their helpful comments. Weiye "Rex" Deng, Dixin Li, Guanzhen Su, and Paco Yip provided generous help in the data collection and categorization process. My colleagues at Washington University, participants at the Quantitative China Studies Seminar at New York University and Harvard Experimental Political Science Conference provided valuable feedback for the paper.

†PhD Candidate, Department of Political Science, Washington University, St Louis, Campus Box 1063, One Brookings Drive, MO 63130, Email: yangzirui@wustl.edu

# Introduction

Recent scholarship claims that public awareness of censorship activities will lead to backlash against the regime and its censorship policy (for an overview, see: Roberts 2020). According to these studies, citizens exposed to censorship events express more anger and anti-regime sentiment (Pan and Siegel 2020; Roberts 2018, 137),[1] discuss and search for more information on the censored topics (Nabi 2014; Pan and Siegel 2020; Roberts 2018, 143), show less support for Internet regulation (Roberts 2018, 144), and even participate in protests against the regime (Boxell and Steinert-Threlkeld 2019). However, in China, where government censorship is massive and unprecedented (Gueorguiev and Malesky 2019; Han 2018; King, Pan and Roberts 2013; Miller 2018), surveys consistently find that Chinese citizens are mostly apathetic toward or even supportive of the regime's censorship policy. For example, Dickson (2016, 71) reports Chinese citizens who have experienced censorship are "rather blasé about it." Roberts (2018, 136) finds the largest proportion of users with censorship experience say "they wouldn't care" about censorship. Wang and Mark (2015) show that a majority (65.6%) of "censorship-aware respondents" are either neutral or supportive of Internet censorship. If awareness of censorship leads to backlash, why do a majority of Chinese citizens display such little resistance to the widespread use of overt censorship?

This puzzle likely arises because the literature that developed the backlash argument primarily focuses on censorship of government criticism and collective action like protests. Although political content has traditionally been understood as the prime targets of censorship (Gueorguiev and Malesky 2019; King, Pan and Roberts 2013, 2014; Miller 2018; Roberts 2018), other forms of censorship, such as prohibitions of pornography (King, Pan and Roberts 2013), Hip-Pop music (Nie 2021), and pop culture movies (Esberg 2020), are also widely used in authoritarian regimes. In this article, I argue that when the range of censored content includes *both politically threatening and seemingly harmless non-political content*, citizens

[1] Pan and Siegel (2020) study the repression of online activists in Saudi Arabia, which is analogous to political censorship.

are less likely to view censorship as suppression of opposition. Rather, censorship becomes viewed as a normal government policy that regulates both political speeches and apolitical content like entertainment, culture, and advertisement. I call this process *normalization of censorship*.

My argument about censorship normalization builds on the psychological theory of desensitization (Bartholow, Bushman and Sestir 2006; Carnagey, Anderson and Bushman 2007; Fanti et al. 2009). Conventionally, citizens react negatively to censorship of political content because it could be seen as a signal that the government has something to hide and is not acting as a faithful agent (Lorentzen 2014; Roberts 2018; Shadmehr and Bernhardt 2015). When both politically threatening messages and non-political content are censored, it dilutes the probability that each censorship event contains valuable political information to discover government wrongdoings (Pan and Siegel 2020; Roberts 2020). In addition to dilution, censoring non-political content also increases citizens' exposure to censorship activities, which further facilitates the normalization process. As individuals are more frequently exposed to censorship activities, they are more likely to view censorship activities as normal events and not react as intensely. As a result of dilution and repeated exposure, citizens' negative reactions toward censorship, such as anger and anti-regime sentiment, should be less likely to occur when censorship is applied to a broader range of content (Wang and Mark 2015).

To support my theory of censorship normalization, I first use observational data to illustrate the possibility that censorship normalization happens in China. I collected and categorized 15,872 censored articles on WeChat from March 2018 to May 2020. Using both human coders and supervised text analysis, I show that collective action, government criticism, and other government-related content only account for around 40% of all censored articles. The majority of censored articles are politically non-threatening and include a wide range of apolitical topics, such as business, entertainment, sexuality, advertisement, local events, traditions, culture, and foreign news.

I then conducted an original survey experiment in China to compare traditional censorship which primarily targets politically threatening messages (government criticism and collective action) with normalized censorship whose targets include both political and non-political content. In the experiment, I randomly expose respondents to varying amounts of censorship of non-political content. Consistent with the normalization theory, respondents exposed to the censorship of both political and non-political content display significantly less backlash and greater support for the censorship policy and the regime, compared with respondents exposed to only censorship of political content.

The results of this study contribute to the understanding of citizens' reactions and resilience to government censorship. It bridges the gap between the seemingly contradictory observations that on one hand, censorship awareness will lead to backlash, and on the other hand, citizens are numb to the massive overt censorship activities in China. By showing the desensitizing effect of large-scale overt censorship, I expand the existing understanding that overt censorship is effective primarily because it creates fear and deters dissent (Roberts 2018, 2020), whereas covert forms of censorship, such as "friction" and "flooding," are more effective in avoiding public backlash (Miller 2018; Roberts 2018, 2020). Overt censorship when applied broadly might be less likely to lead to backlash as well.

A broader implication of these findings concerns the toolkits of authoritarian governments. A wealth of scholarship has investigated how authoritarian governments persuade people through propaganda, silence dissent through censorship, and prevent uprisings through repression (Arendt 1973; Cantoni et al. 2017; Chen and Xu 2017; Geddes and Zaller 1989; Guriev and Treisman 2015; Huang 2018; Shadmehr and Bernhardt 2015; Svolik 2012). My study suggests that there might be an additional and powerful tool of the autocrats: desensitizing the citizens to coercion by normalizing the coercive policy. Such an approach is potentially more difficult to undermine because citizens unconsciously accept the coercive policy as normal.

# Normalization of Censorship: A Theory

The central argument of this paper is that, when the targets of censorship include both politically threatening and seemingly harmless non-political content, it normalizes censorship policy and desensitizes citizens to censorship activities. As a result, backlash against both the censorship apparatus and the regime is less likely to happen. Such normalizing effects are primarily achieved by diluting the proportion of politically threatening content among censorship targets. In addition to dilution, censoring non-political content also increases citizens' exposure to censorship activities, which further facilitates the normalization process.

## Diluting the Proportion of Politically Threatening Content

Backlash to censorship happens when citizens are aware of censorship activities and care about what has been censored (Chen and Yang 2019; Roberts 2020). The extent to which citizens care about censorship activities depends on the nature of the censored content. Conventionally, the government tends to target the most politically threatening information, such as messages with collective action potential and government criticism (Lorentzen 2014; King, Pan and Roberts 2013, 2014; Shadmehr and Bernhardt 2015). Hence, censorship could be seen as a signal that the regime has something to hide and is not acting as a faithful agent for the citizens (Lorentzen 2014; Roberts 2018; Shadmehr and Bernhardt 2015). It indicates abnormality and potential government wrongdoings. As a result, citizens will pay even closer attention to the censored information to find out what has been hidden from them. Such an effect is called the Streisand effect: the act of censorship drawing even more attention to the event that the government initially tried to cover up (Roberts 2020). Given the difficulty of completely covering up information on the Internet (Roberts 2018), once the citizens uncover the censored information, the anger toward the government will be magnified. Consistent with this logic, several recent studies have found evidence of backlash against censorship of political content (Pan and Siegel 2020; Roberts 2018, 2020).

For such backlash to occur, however, it is critical that citizens believe censorship is abnormal and censored information is valuable for discovering government wrongdoings. If citizens view censorship as normal, they are less likely to pay attention to censorship events in the first place. Therefore, a Streisand effect is less likely to occur.

According to psychological research of desensitization, when subjects' categorization and expectation of a stimulus are shifted from negative to neutral (or even positive), they will be less sensitive to the stimulus and their negative reactions to the stimulus will be diminished (Carnagey, Anderson and Bushman 2007; Efran and Marcia 1967; Goldfried 1971; Marcia, Rubin and Efran 1969). For example, playing violent video games will expose subjects to initially negative stimuli (i.e. violence) in a positive emotional context. As a result, subjects will change their normative evaluation of violence and decrease their attention to violent events (Carnagey, Anderson and Bushman 2007). Similarly, if citizens are exposed to censorship in a neutral or positive context, they will view censorship as normal and less likely to pay attention to censorship events. As a consequence, subsequent backlash against censorship is less likely to happen.

Under what conditions will citizens be exposed to censorship in a neutral or even positive context? Direct interaction with censorship, such as having one's own message censored, is likely to be negative and cause backlash, because their own rights of free speech are directly infringed. In contrast, indirect interaction with censorship, such as observing a web page blocked or discussing censorship online, might be less negative and might not lead to backlash. When users observe a deleted post, they rely on the remaining information such as the title of the post, the author of the post, and other users' comments and reactions under the post to update their beliefs about censorship.[2] If censored content is always politically sensitive, users are likely to associate censorship with negative news of the government. In contrast, censorship of non-politically sensitive content dilutes that negative image. When

[2]How much information remains after censorship depends on individual websites and social media platforms. On WeChat, if an article is deleted, users can still see the title of it. Only if they try to click on the link to access the full article will they find out the content is censored. Sina Weibo has more varieties of censorship. See Miller (2018) for more information about Sina Weibo.

encountering censorship of non-political content, users will update their beliefs and be less likely to associate censorship with hiding politically valuable information or silencing political dissent in the future.

Moreover, users often talk about censorship in online conversations in neutral and apolitical contexts (Han 2018). For instance, in February 2020, the fans of Xiao Zhan – a Chinese entertainment star – and other fans engaged in an argument that frequently involved demanding censorship of the other side's posts. Such conversation implies that censorship can be applied to non-political posts on entertainment stars. People observing it might be less likely to think that censorship is negative and associated with political repression.[3]

According to existing surveys, only 9% of the respondents experienced censorship directly (Dickson 2016), even though 69.5% of the respondents are aware of government censorship (Wang and Mark 2015). Hence, direct interaction is relatively rare and most citizens form beliefs about censorship via indirectly interacting with censorship. Therefore, the range of censored content has a strong influence on citizens' beliefs toward censorship. By including seemingly harmless non-political topics among censorship targets, the value of censorship activities in signaling government wrongdoing is significantly diluted. In response, citizens update their beliefs about censored information and are less likely to display as much backlash when they encounter censorship (Hill 2017; Wang and Mark 2015). Furthermore, if unpopular content, such as pop culture in the eye of conservative citizens, is also censored, citizens might even think positively about these specific censorship activities and be more supportive of the censorship apparatus in general (Esberg 2020).

In short, including non-political content among censorship targets dilutes the proportion of politically threatening content and changes citizens' belief that censorship is abnormal and hiding politically valuable information such as government wrongdoings. As a result, the initial negative reactions to censorship of political content should be less likely to occur.

[3]I treat actively demanding censorship as a sign of normalization because it shows that Internet users no longer regard censorship as a taboo or a sensitive topic, but rather a normal policy that could be utilized. Nevertheless, actively demanding censorship is a special and interesting downstream behavior of the normalization process. Future research should explore the consequences of demanding censorship.

## Increasing Citizens' Exposure to Censorship

In addition to dilution, including non-political content among censorship targets increases the frequency of citizens' exposure to censorship, which further facilitates the normalization process. As stated above, the belief that censorship is abnormal is critical for backlash to occur. If the chance of encountering censorship increases in citizens' daily lives, it is more likely for them to view censorship as normal and not pay too much attention to it.

A deeper look into the psychological mechanism suggests that such a desensitizing effect is due to the blunted reactions after repeated exposure to similar stimuli. Initially, a negative stimulus, such as violence or repression, arouses cognitive, physiological, and emotional responses (Anderson et al. 2010; Bartholow, Bushman and Sestir 2006; Carnagey, Anderson and Bushman 2007). Repeated exposure to the same stimulus, even over a short period of time, leads to blunted evaluative categorization and elimination of physiological and emotional reactions (Bartholow, Bushman and Sestir 2006; Carnagey, Anderson and Bushman 2007; Fanti et al. 2009). Similarly, although the initial exposure to censorship might arouse intense cognitive and emotional reactions, such as anger and resentment, such cognitive and emotional responses should be less likely to occur as individuals are more frequently exposed to censorship activities. As a result, citizens are more likely to regard censorship as normal.

The normalizing effect of repeated exposures further facilitates indirect interactions with censorship. Because citizens regard censorship as normal and common in daily life, they would not deliberately avoid it in conversation like they do with other sensitive topics. The example above about the online argument involving Xiao Zhan's fans illustrates how censorship could be a normal topic that is frequently mentioned in online conversations. The effect of repeated exposure becomes a positive circle that reinforces the belief that censorship is normal.[4]

[4]It would be interesting to consider whether there is a limit to how much information can be censored until the normalization effect disappears. Existing studies do show that censorship of some non-political content, such as banning certain social media altogether, can lead to backlash (Hobbs and Roberts 2018). Hence, there is likely to be a limit beyond which backlash occurs. In practice, although the government has been expanding the range and scale of censorship, it still constitutes only a tiny fraction of the enormous

### Empirical Expectations

To summarize the empirical expectations of the theoretical arguments laid out above, I hypothesize that *citizens exposed to normalized censorship, which targets both politically threatening and seemingly harmless non-political content, will display less backlash against both the censorship policy (H1) and the regime (H2), compared with citizens exposed to traditional censorship which only targets politically threatening content.*

Before I test the two main hypotheses, however, I need to illustrate the possibility that censorship normalization happens in China. One observable implication of the normalization strategy is that non-political content accounts for a large proportion of all censored content, i.e., Pr(Non-Political Content|All Censored Content) is high. To be clear, this is not to say that non-political content is more likely to be censored than political content, i.e., Pr(Censorship|Non-Political Content) > Pr(Censorship|Political Content), or vice versa. I do not test which category is more likely to be censored. Instead, I aim to verify that censorship of non-political content happens on a substantial scale. In the following sections, I first present an observational study that uses text analysis to identify the nature of censored content and then present experimental evidence supporting my main hypotheses.

## The Development of Censorship in China

I test my theory of censorship normalization in mainland China. Although online censorship conducted by the government or social media companies occurs in almost every country, especially in authoritarian regimes, the range and scale of government censorship in China is by far the largest (King, Pan and Roberts 2013; Roberts 2018). Earlier studies have found that the Chinese government primarily targets content with collective action potentials while allowing most other online expressions (King, Pan and Roberts 2013, 2014). However, in

amount of data online (Roberts 2018). As such, there is likely to be enough leeway for the government to conduct censorship without causing backlash. In the context of the current study (i.e. not blocking the Internet altogether), it is hard to say exactly how much information is too much for a ceiling effect to occur, but it is an interesting question for future research.

recent years, overt censorship activities in China have become more aggressive (Freedom House 2019). In addition to banning more political content than previously understood (Gueorguiev and Malesky 2019; Miller 2018), many seemingly harmless posts that include sensationalism, speculation, and tabloid gossip are also censored (Cairns and Carlson 2016; Han 2018; Huang 2017; Ng 2015). Once-tolerated platforms focused on apolitical topics, including entertainment and dating applications, faced new restrictions (Freedom House 2019). Information and discussions on subjects like the economy that have traditionally been given freer rein became more systematically censored (Tai and Fu 2020).

Meanwhile, the Chinese government and social media companies have become more explicit on their intention to regulate online expression and use “inappropriate” content to justify its censorship policy. In 2014, the Chinese government established the Cyberspace Administration of China (CAC) to centralize the administrative power and reduce agency loss that used to provide leeway for social media companies to partially resist government censorship efforts (Han 2018; Miller 2018). Since its establishment, the CAC has been regularly publishing its censorship activities on its official website, which often receives media attention. Social media companies have also started to publicly acknowledge and justify their censorship activities. In April 2020, WeChat, the largest social media platform in China, publicized its censorship of more than 15,000 articles on its platform (The Paper 2020).

More importantly, the justifications for these publicized censorship activities are mainly apolitical which diverts sharply from traditional threats of repression or showcase of regime strength. When mentioning censorship in public, the government often emphasizes the dark side of the Internet and highlights the need for Internet regulations. For example, the Party General Secretary Xi Jinping once noted in a public speech:

> “No one wants to live in a [cyberspace] full of falsehood, fraud, attack, abuse, terror, pornography, and violence. The Internet is not outside the law. Tightening Internet regulation and strengthening Internet governance are the responsibility of the government to the people and the society.” (Xinhua 2018)

Other officials, such as the former director of the CAC, also publicly referred to censorship as "management" of the Internet, and use "inappropriate" content as a justification for government censorship activities (China Youth Daily 2015). Consistent with such a narrative, WeChat framed its censorship as countering false information (The Paper 2020). This further shows to the public that the focus of censorship is not political dissent but "inappropriate" content such as false information.

Combining the different pieces of anecdotes above, although the government censorship activities have been expanding rapidly in recent years, the image of the censorship policy in China has become more benevolent and less politically repressive.

## The Nature of Censored Content: Text Analysis

To better illustrate the possibility that censorship normalization happens in China, I use text analysis to more rigorously show that the censorship of non-political content occurs on a substantial scale. I collect censored articles from March 2018 to May 2020 on WeChat, the largest social media platform in China. I then classify the censored articles into nine different topic categories, including three political and six non-political categories. The main outcome of interest is the proportion of censored articles by topic category. To ensure the reliability of the categorization process, I use two human coders, multiple text analysis models, and cross-validations.

### Data Source

My observational study relies on the WeChatScope data of censored articles on WeChat public accounts. WeChatScope is a website created by a research team at the University of Hong Kong (Tai and Fu 2020), which monitors 4000 WeChat public accounts in real-time. These public accounts are analogous to Facebook public pages or Telegram public groups. Because these public accounts might have a large number of subscribers and therefore be influential, they are the prime target of WeChat censorship. The dataset contains 15,872 articles spanning from March 2018 to May 2020.

WeChat is an ideal platform for analyzing the implication of government censorship on citizens' attitudes because it is the single most popular social media in Mainland China. By 2020, it has over 1.2 billion daily active users, and a 78% penetration rate (Iqbal 2020). As such, WeChat censorship influences a large proportion of the Chinese population and most Chinese are likely to form their beliefs about censorship via WeChat.

Although censorship practice on WeChat might not be identical to other social media platforms or government censorship instructions due to principal-agent problems (Miller 2018), the inconsistency has reduced significantly in recent years. Since the establishment of the CAC in 2014, the government has gained considerable leverage over social media platforms. Platforms that fail to enforce government censorship orders are more seriously punished by the CAC than before its establishment. Overall, there is no evidence suggesting either atypical censorship practice on WeChat or bias against political content.

Using the WeChatScope data provides a hard case to illustrate the existence of large-scale censorship of non-political content. Because the WeChatScope project is designed to capture government censorship as conventionally understood, political accounts are over-represented in the sample (Tai and Fu 2020).[5] Therefore, even if the selected public accounts might not be representative of all public accounts, the bias is likely to be in the opposite direction of my theoretical expectations.

One caveat to the WeChatScope data is that, like most quantitative censorship data, it only includes *post hoc* censorship.[6] The articles need to be published on WeChat before they could be manually censored and recorded in the database. However, an article has to pass *ex ante* censorship barriers such as "keyword blocking" and "The Great Firewall of China"

[5]The WeChatScope project primarily includes accounts related to social and political news or commentary. It also samples influential public accounts including (a) public accounts for the government and the Communist Party; (b) high ranked accounts; and (c) accounts with article links posted on a major discussion forum or indexed by Baidu search engine (Tai and Fu 2020).

[6]I thank the Anonymous Reviewer for pointing out this caveat.

before it could be posted online (King, Pan and Roberts 2013). According to previous analyses of taboo keywords, *ex ante* censorship is focused overwhelmingly on political topics (Han 2018; Ng 2015). Thus, political content might have a higher bar for publication than non-political content. Nonetheless, citizens have creative ways to bypass keyword blocking and *post hoc* censorship is the most extensive form of censorship (Han 2018; King, Pan and Roberts 2013). As such, I focus on *post hoc* censorship in this analysis.

## Categorization of Censored Articles

I aim to examine the proportion of censored articles by topic category. As such, categorizing the censored articles is central to this analysis. In total, I keep track of nine different topic categories. In addition to three political categories: (1) collective action, (2) government criticism, and (3) government non-criticism, I also include six non-political categories: (1) business & economy, (2) entertainment & sexuality, (3) advertisement, (4) local events, traditions, cultures, (5) foreign news, and (6) others.

The categorization process and coding rubrics mainly follow Miller (2018), because Miller (2018) provides the most detailed, reliable, and up-to-date categorization of censored content in China. One important difference from Miller (2018) is that the nine categories are mutually exclusive. A similar strategy is employed by King, Pan and Roberts (2013). Having mutually exclusive categories simplifies the categorization process as well as the interpretation of the results. In practice, the nine categories are coded sequentially with political categories coded first. Specifically, an article will first be considered if it belongs to the collective action category. If yes, then the coding process ends. If not, the article will then be considered if it belongs to the government criticism category, and then the government non-criticism category, and so on. Such a coding process ensures that collective action and government criticism will not be underestimated in the analysis. The detailed explanation of each topic category and coding process can be found in Online Appendix D.

To categorize more than 15,000 articles, I first hand-label a training set of 2,500 articles and then use supervised text analysis to predict the categories of the remaining articles. Two coders have coded the training set independently. The Cohen's $\kappa$ between the two coders is 0.80, higher than the commonly applied criteria of 0.70 for inter-coder reliability tests.

To predict the remaining 13,000 unlabeled data, I first train a multinomial logistic regression model with ridge estimator using the 2,500 labeled data (Hoerl and Kennard 1970; Le Cessie and Van Houwelingen 1992). Then, I use the model as well as the text corpus to predict the unlabeled data. The details of the model can be found in Online Appendix E.

## Results

Table 1 reports the predicted proportion of censored articles by topic category. Consistent with the empirical expectations, a substantial proportion of censored articles are unrelated to politically threatening topics. As shown in table 1, collective action and government-related articles only account for 39.34% of all censored articles, whereas non-politically threatening content consists of the majority (60.67%) of censored articles on WeChat. However, there is not a clear pattern among these non-political articles. Around 14.43% of the articles are categorized as business and economy. Another 20.32% belongs to the entertainment and sexuality category. The advertisement category, the local events, traditions and cultures category, and the foreign news category account for around 7.4%, 12.28%, and 3.38% of the censored articles respectively. The remaining 2.86% of the articles do not fit into any specified topic category and therefore fall into the residual group of articles.

The pattern that non-politically threatening content constitutes a majority of the censored articles is also mostly consistent over time. Figure 1 shows the time series of the proportion of collective action, government criticism, government non-criticism, and other non-politically threatening topic categories. Even during the COVID-19 outbreak in early 2020 when government suppression of online criticisms was most active, non-politically threatening content still accounts for around 48% of all censored articles.[7]

[7] The sudden surge in political suppression of online criticisms during the COVID-19 outbreak did trigger public outrage, which was followed by an online movement demanding freedom of speech. This illustrates

Table 1: Predicted Proportion of Censored Articles by Topic Category

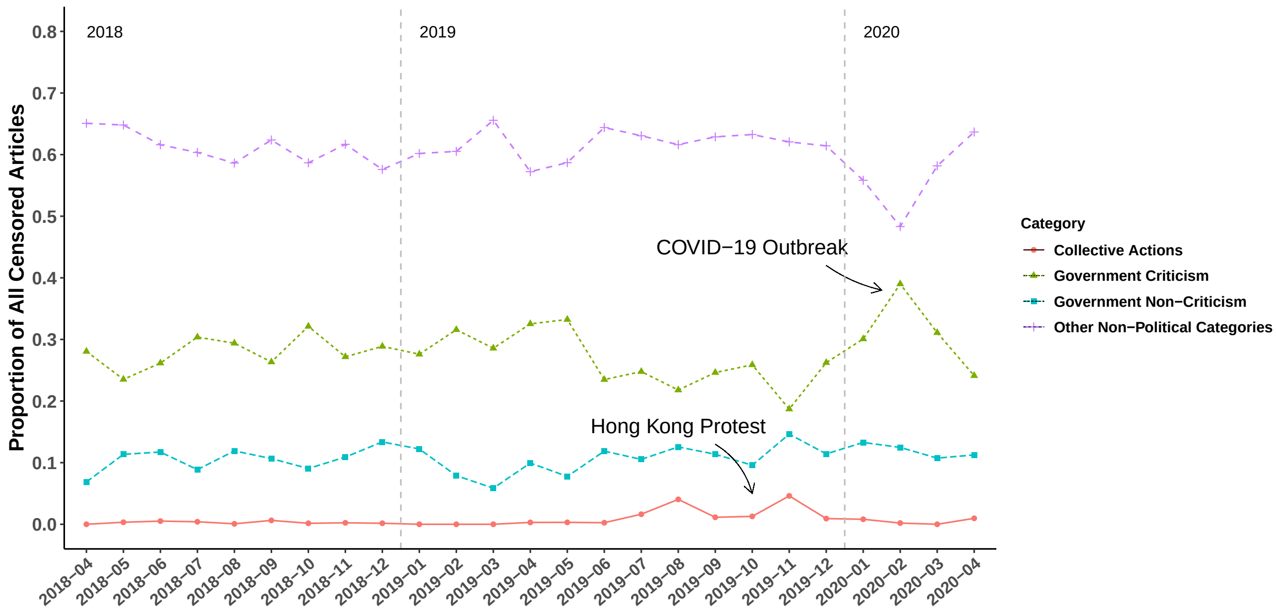
| General Category | Percentage | Specific Topic Category | Percentage |
|---|---|---|---|
| Political | 39.34% | Collective Action | 0.67% |
| | | Government Criticism | 27.94% |
| | | Government (Non-Criticism) | 10.73% |
| Non-Political | 60.67% | Business & Economy | 14.43% |
| | | Entertainment & Sexuality | 20.32% |
| | | Advertisement | 7.40% |
| | | Local Events, Traditions, Cultures | 12.28% |
| | | Foreign News | 3.38% |
| | | Others | 2.86% |

*Notes:* The total number of articles is 15,872. A training set of 2,500 articles is labeled by two human-coders independently. In-sample cross-validation suggests that government criticism category is over-estimated by the model, whereas most other categories are under-estimated. In the training set, both coders find around 30% of the articles are political.

Alternative text analysis models as well as a three-fold cross-validation are used to test the robustness of the results. The predictions are mostly consistent with the multinomial model, showing the existence of large-scale censorship of non-political content. More details of the alternative models and cross-validation can be found in Online Appendix E and F. The implications of the empirical results are two-fold. First, the results from WeChat censorship echo the narrative in the previous section showing that the Chinese government has spent considerable effort in censoring non-politically threatening content while downplaying suppression of political discussion. The fact that collective action and government-related articles account for a minority of all censored content strengthens the government's claim that censorship is normal and benevolent. Besides, the fact that there is not a clear pattern for non-political content implies that the government censorship policy is broad and comprehensive rather than narrow and focused on one specific type of non-political content such as popular culture (Esberg 2020) or pornography (King, Pan and Roberts 2013). This gives the government leeway for including politically threatening content in the broadly targeted

that times of crisis might be a potential pitfall of the normalization strategy. The COVID crisis also shows that, if a crisis persists despite censorship, backlash is more likely to occur even if the government tries to normalize censorship. Future studies should further theorize censorship in the time of crisis.

Figure 1: Time Series of All Censored Articles by Topic Category




*Notes:* Other non-political categories include: Business & Economy, Entertainment & Sexuality, Advertisement, Local Events & Cultures & Traditions, Foreign News, and Others.

censorship activities without drawing too much attention to the political content.

Second, the results imply that Chinese citizens are more likely to encounter censorship of non-political content than censorship of government criticism and collective action. As a result, their expectations of censorship outcomes are likely to be shaped by these experiences of censorship of non-political content, which provides favorable conditions for the desensitization process to take place.

# Normalizing Effect of Censorship: Survey Experiment

The previous section illustrates the possibility that censorship normalization happens by showing the existence of large-scale censorship of non-political content. The result verifies the validity of the two main hypotheses. To test the two main hypotheses, I experimentally manipulate the topics of censorship targets and the frequency of censorship activities that participants are exposed to. I then use their *ex post* evaluations of the censorship policy and the regime to measure the level of backlash induced by censorship.

## Participants

The survey experiment was conducted in December 2020 in mainland China.[8] The 612 participants were recruited from a Chinese online survey platform and then directed to an American-based website, Qualtrics, where they completed the survey anonymously. Previous research has shown that using online platforms such as Amazon Mechanical Turk is a reliable way to recruit participants for survey experiments (Mullinix et al. 2015), and Chinese online platforms might even be better than American online platforms because political opinion surveys are relatively rare in China and participants are less likely to be professional political survey takers (Huang 2018; Huang and Yeh 2019).[9]

The sample in this study covers a wide range of socioeconomic backgrounds and is demographically similar to the Chinese Internet population in terms of gender, urban/rural, and regional distributions. However, like many other online surveys in China, the sample is richer and better-educated than the general Internet population (Huang 2018). To address these concerns, I conduct interaction analyses and find no significant heterogeneous treatment effect with respect to age, income, and education.[10] Hence, it is unlikely that this specific sample causes bias in the estimates. More detailed descriptive statistics of this sample are available in Online Appendix A.

[8]This study is pre-registered on the Open Science Framework via EGAP registry. EGAP Registration ID: 20201204AB.

[9]At the end of the survey, respondents were allowed to leave comments on the survey questions. The vast majority of the comments were positive and many said "[they] hope there could be more these kinds of surveys in China." This further shows that participants were excited rather than scared by political surveys. Due to the tension between China and the U.S. during the period of the study, Chinese respondents may respond differently to a survey from an American-based institution than to a Chinese survey company. Hence, any information in the survey that could possibly reveal the fact that this is a survey from a foreign institution, including logo, header, and country-specific URLs, was removed or replaced. No respondent left any comment about the U.S. or any other countries at the end of the survey.

[10]See Online Appendix B.3.

## Experimental Design

The purpose of the experiment is to compare traditional censorship which primarily targets politically threatening messages (government criticism and collective action) with normalized censorship which targets both politically threatening and seemingly harmless non-political content. The experiment includes three components. First, participants answered pre-treatment questions about their socioeconomic backgrounds, political interests, economic ideology, and exposure to social media. Second, participants were randomly assigned to either a control group where they would be exposed to only censorship of government criticism and collective action,[11] or a treatment group where they would be exposed to censorship of *both political and non-political content*. Finally, they answered post-treatment questions about their attitudes toward the censorship policy and the regime.

To expose participants to censorship, I asked respondents to read ten snippets of WeChat articles, presented one at a time with only the title and first few lines.[12] Some of the snippets were labeled as censored by WeChat. These censorship labels primed the respondents that certain articles were censored by WeChat. No further information was provided to the respondents. Figure 2 shows one of the snippets with and without a censorship label.[13] Among the ten snippets, six were about non-political topics, whereas four were about government criticism and collective action. The order of the snippets was randomized.

I validated the appropriateness of my choice of articles in two ways. First, I consulted a panel of China scholars about the snippets. None of them thought that any of the snippets

[11] Encountering censorship of only political content is still a plausible scenario even in a context where normalization is already taking place in China. This is because normalization only increases the probability of encountering censorship of non-political content. It is still possible for citizens to only encounter censorship of political content.

[12] The snippets are screenshots of real articles censored by WeChat. They only include the first couple of lines and do not reveal the full content of the articles. I selected these snippets from the WeChatScope dataset used in the observational study. The selection process was systematic. I first randomly drew twenty articles from the dataset and then selected ten out of the twenty articles which I deemed as appropriate for the experiment.

[13] In reality, if an article is censored by WeChat, users can only see the title of that article but not the first couple of lines. In this experiment, to ensure the only difference between the treatment and control groups is the censorship label, I use the same snippets for both groups.

Figure 2: Example Article Snippet with and without Censorship Label (The Red Exclamation Mark) in the Experiment


*Notes:* The censorship label reads: This article was blocked by WeChat due to violation of Internet law.

were absurd, fraudulent, or otherwise unreasonable. Second, I also asked the respondents about their interest in reading the full article after they read each snippet. If a snippet is particularly sensationalist, fraudulent, or "deserved to be censored," then this would be reflected in an unusually high or low level of interest from the average respondent. None of the snippets stands out as exceptionally high or low in terms of the level of interest. Hence, the experimental results are not driven by "inappropriate" or unreasonable choice of articles.

In the control group, three out of the four political snippets were labeled as censored by WeChat, whereas none of the non-political snippets were labeled. This primed the respondents in the control group that censorship primarily targets politically threatening content. In the treatment group, three out of the six non-political snippets were labeled as censored by WeChat in addition to the three political snippets in the control group. In other words, there were a total of six snippets that were labeled in the treatment group and three in the control. This primed the respondents in the treatment group that both political and non-political content could be censored. This also means that respondents in the treatment group are exposed to twice as many censored snippets as those in the control group. Hence, the treatment reflects both dilution and repeated exposure laid out in the theory.

Table 2 summarizes when and where the censorship label occurred in the treatment and control groups. The difference between the treatment and control groups is that respondents in the treatment group were exposed to additional censorship of three non-political snippets. Labeled snippets remained constant across subjects in respective groups, though as mentioned, their orders were randomized (i.e. labeled snippets might occur at any position). The details of the ten article snippets can be found in Online Appendix C.

Table 2: Explanation of Treatment

| Article Snippets | Government or Collective Action | Control Group (Traditional Censorship) | Treatment Group (Normalized Censorship) |
|---|---|---|---|
| Snippet 1 | No | | Censorship Label |
| Snippet 2 | No | | Censorship Label |
| Snippet 3 | No | | Censorship Label |
| Snippet 4 | No | | |
| Snippet 5 | No | | |
| Snippet 6 | No | | |
| Snippet 7 | Yes | Censorship Label | Censorship Label |
| Snippet 8 | Yes | Censorship Label | Censorship Label |
| Snippet 9 | Yes | Censorship Label | Censorship Label |
| Snippet 10 | Yes | | |

*Notes:* The difference between the treatment and control groups is that the treatment group were exposed to additional censorship of three non-political article snippets. Column 1 indicates whether each of the snippets is political or non-political. Both groups read all ten snippets. In the control group, snippet 7, 8, and 9 (all political) were labeled as censored, whereas in the treatment group, snippet 1, 2, 3, 7, 8, and 9 (three political and three non-political) were labeled as censored. The order of the snippets was randomized in both groups.

## Measurement

The main dependent variables of the study are backlash against the censorship policy and backlash against the regime. To measure backlash against the censorship policy, I first asked a straightforward question about respondents' *support for government control of the Internet*. However, asking straightforward questions in authoritarian regimes might induce social desirability bias and potentially underestimate treatment effects. Hence, I asked two additional questions on censorship policy. The second question about policy backlash asks respondents whether they think *government control of the Internet is normal*. This question also directly tests the normalization theory that claims citizens will accept a coercive policy as normal. Finally, I asked about their *perception of censorship volume*. If respondents perceive censorship volume as too high, they might think that current censorship is too excessive and hence prefer less stringent censorship, and vice versa. Despite encountering twice as many censorship labels, I expect the treatment group to express a lower perception of censorship volume and prefer more censorship than the control group.

To measure backlash against the regime, I directly borrowed questions measuring *assessment of the government*, *overall satisfaction of China*, and *willingness to participate in protests* from Huang (2018). Similar to the policy backlash questions, the questions about regime backlash include both straightforward measures of regime support and a more indirect measure of overall satisfaction. Higher support and overall satisfaction indicate a lower level of regime backlash. For assessment of the government, I asked separately about central and local governments. Because discussing local government is less sensitive than central government, it alleviates potential social desirability bias and ceiling effects problems. For the last question, I expect respondents in the treatment group to be less willing to participate in protests. However, this is not because of fear and outright intimidation from the regime (Huang 2018; Young 2019). Instead, citizens become more supportive of the regime and less willing to protest due to a sincere belief that censorship is normal and the government is for the people. Table 3 summarizes the wording of each question and the expected direction of treatment effects.

Because I use multiple items to measure the same dependent variables, it might introduce

Table 3: Measurement of Main Outcome Variables

| Hypothesis | Survey Items | Expectation |
|---|---|---|
| Policy Backlash | Do you agree or disagree with the following statement: The government should actively control the Internet and remove content that it deems inappropriate. | + |
| | Do you agree or disagree with the following statement: For most of the time, the current government's control of the Internet are normal and justified. | + |
| | Do you think the level of Internet control in China is too high, too low, or about right? | − |
| Regime Backlash | Do you agree or disagree with the following statement: Our central government always works for the people and serving their needs. | + |
| | Do you agree or disagree with the following statement: Our local government always works for the people and serving their needs. | + |
| | How satisfied are you with the overall situation in China right now? | + |
| | If there is a gathering of people or "collective walk" over the work of the government in your area, would you participate in it? | − |

*Notes*: All individual survey items were measured on a five-point scale, with 5 indicating "Strongly Agree," "Very Satisfied," or "Very Likely." For question 3 of policy backlash, 5 indicates "Way too high," 3 indicates "About right," and 1 indicates "Way too low." The order of the questions was randomized.

bias due to multiple hypothesis testing. I address this concern by presenting $p$-values that are adjusted using the most conservative Bonferroni correction in which the $p$-values are multiplied by the number of tests.

I use nine covariates to check the balance between the treatment and control groups. They are also included in the regression analyses. Among the nine covariates, four are demographic variables, including *Age*, *Income*, *Female*, and *Education*. The remaining five covariates are *Pro-Market Ideology*, *Party Member*, *Political Interests*, *Social Media Usage*, and *VPN Usage*, which measure participants' political predispositions and Internet usage. All covariates are balanced between the treatment and control groups and none of them significantly predicts treatment assignment. Hence, the randomization process is successful. Details of the balance table are reported in Online Appendix A.

## Results

Table 4 reports the two-tailed $t$-test results comparing the control and treatment groups on individual survey items. For attitudes toward censorship policy, the treatment significantly increases respondents' support for government control of the Internet ($\beta = 0.264$, $p < 0.01$) and their beliefs that censorship is normal ($\beta = 0.243$, $p < 0.01$). Both results indicate that the range of censorship targets matters for public reactions toward censorship. Additional exposure to censorship of non-political content reduces respondents' backlash against censorship policy even if they are also exposed to censorship of political content. By contrast, the treatment effect on respondents' perception of censorship volume is insignificant but in the correct direction ($\beta = -0.105$, $p = 0.34$). The insignificant result is perhaps because responses concentrated on the median choice – "About Right" – whereas few respondents chose "Too High" or "Too Low."

Table 4: Two-Tailed $t$-tests Comparing Control and Treatment Groups on Individual Survey Items

| Hypothesis | Survey Items (Theoretical Expectations) | Control (n = 307) | Treatment (n = 305) | Mean Diff ($T - C$) | $t$-stats | Adjusted $p$-value |
|---|---|---|---|---|---|---|
| Policy Backlash | Support for Gov't Control (+) | 3.353 (0.191) | 3.617 (0.207) | 0.264 (0.082) | −3.201 | 0.004 |
| | Censorship is Normal (+) | 3.681 (0.210) | 3.924 (0.225) | 0.243 (0.076) | −3.209 | 0.004 |
| | Perceived Censorship Volume (−) | 2.940 (0.168) | 2.836 (0.162) | −0.105 (0.066) | 1.584 | 0.341 |
| Regime Backlash | Assessment of Central Gov't (+) | 3.957 (0.226) | 4.193 (0.240) | 0.236 (0.071) | −3.316 | 0.004 |
| | Assessment of Local Gov't (+) | 3.686 (0.210) | 3.973 (0.228) | 0.288 (0.080) | −3.582 | 0.001 |
| | Overall Satisfaction of China (+) | 3.860 (0.220) | 4.089 (0.234) | 0.229 (0.073) | −3.132 | 0.007 |
| | Willingness to Protest (−) | 2.895 (0.165) | 2.627 (0.150) | −0.268 (0.102) | 2.638 | 0.034 |

*Notes*: All individual survey items were measured on a five-point scale. Bonferroni corrections were used to correct the $p$-value for multiple hypothesis testing.

For attitudes toward the regime, exposing respondents to censorship of both political and non-political content increases their support for both central ($\beta = 0.236$, $p < 0.01$) and local governments ($\beta = 0.288$, $p < 0.01$), their overall satisfaction of China ($\beta = 0.229$, $p < 0.01$), and decreases their willingness to protest ($\beta = -0.268$, $p < 0.05$). The effect sizes are large given the fact that baseline support for the regime is already high in the control group. The average support for the central government in the control group is 3.96 out of 5, and support for local government in the control group is 3.69 out of 5. As such, the effect sizes of 0.236 and 0.288 are both substantial increases in regime support and substantial decreases in regime backlash.[14]

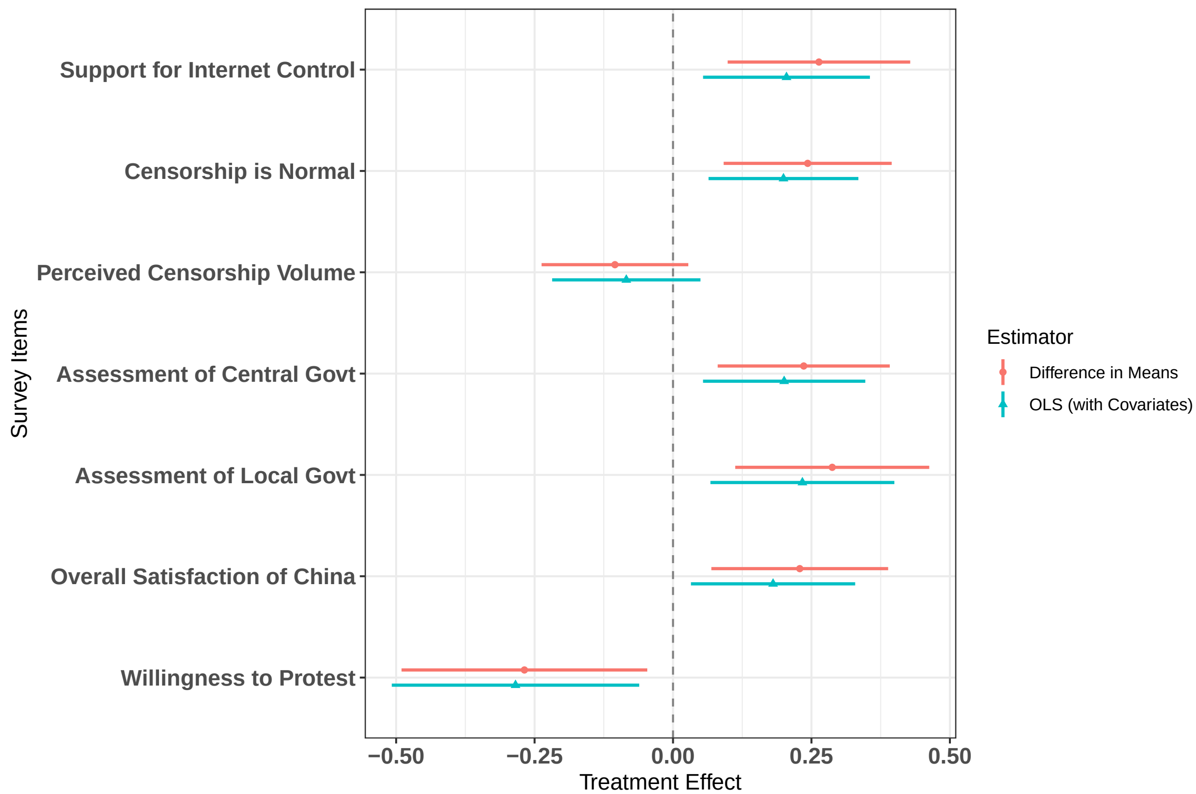
A noteworthy point is that respondents are both more supportive of the government and less willing to participate in protests. Hence, it differs from traditional threats of repression or showcase of regime strength which would lead to lower regime support and lower willingness to protest (Huang 2018; Wang 2019). It is not the case that participants in the treatment group are less willing to protest against the regime due to fear of repression. Instead, it is the result of the sincere belief that censorship is normal and the government is for the people, which is indicated by an increase in policy and regime support.

Figure 3 reports additional results of OLS regression models with control for covariates. The 95% confidence intervals are adjusted for multiple hypothesis testing by Bonferroni corrections. The regression results are consistent with the $t$-tests and support both the policy backlash hypothesis and the regime backlash hypothesis.

The implications of the experimental results are three-fold. First, the results confirm that expanding the range of censorship targets beyond politically threatening content to seemingly harmless non-political content significantly reduces the likelihood of backlash induced by censorship of political content. The Chinese government's effort in expanding and emphasizing the non-political targets of its censorship activities seems to be a plausible explanation of why most Chinese citizens are apathetic toward or supportive of government censorship. When exposed to censorship of non-political content, citizens update their beliefs

[14]The magnitude of the treatment effects are similar to the survey experiment in Huang (2018), which tests the effect of "hard" propaganda on regime support.

Figure 3: Treatment Effect of Additional Censorship of Non-Political Content on Individual Survey Items with 95% Confidence Interval ($t$-tests & OLS Regressions with Covariates Control)




*Notes:* All individual survey items were measured on a five-point scale. Nine covariates were included in the OLS regression models. All 95% confidence intervals were adjusted for multiple hypothesis testing by Bonferroni corrections.

about censorship and are more likely to view censorship as normal rather than repressive. As a result, support for both censorship and the regime increases.

Second, the results show that repressive policies like censorship might be popular and can increase regime support. This suggests that normalization is an independent government strategy different from persuasion, which highlights the achievements of the government, and repression, which showcases regime strength and deters dissent.

Third, the results imply that combating censorship requires not only exposure to censorship activities but also revealing the repressive nature of censorship. Currently, increasing censorship awareness is regarded as an important policy suggestion to counter government censorship (Roberts 2020). However, in this experiment, increasing exposure to censorship activities leads to higher support for censorship and the regime. Hence, revealing the repressive nature, rather than increasing censorship awareness, should be a more effective way to counter government censorship. Combining the results from the survey experiment and the observational study, I show that large-scale censorship of non-political content both exists in China and can significantly reduce the likelihood of backlash against the regime and its censorship policy.

## Discussion and Limitations

Before concluding, let me also discuss some potential limitations to both the normalization theory and the empirical analysis.[15] First, theoretically, does censoring *any* non-political content have the normalizing effect? Will censoring non-political content ever backfire? It is well established in the political science literature that people care more about non-political issues like sports and entertainment than political issues (Carpini and Keeter 1996; Roberts 2020). Therefore, it is logical to argue that censoring some of the popular non-political topics might lead to backlash as well. Consistent with this logic, Hobbs and Roberts (2018) find that the Chinese government's blocking of Instagram, a primarily non-political social media platform, inspired millions of Chinese users to acquire virtual private networks to bypass censorship and subsequently get access to forbidden political information. This suggests that the censorship of non-political content at a large scale is not without possible public backlash. Although, as explained in the experimental design section, the snippets used in the experiment are normal non-political articles censored by WeChat, there may still be strategic considerations in choosing *which* non-political topics to censor. This remains an important topic for future research to explore.

In addition to the theoretical limitation, there are two empirical limitations. First, despite showing evidence of some observable implications of the normalization theory, both the

[15] I thank the Anonymous Reviewers for pointing out these limitations.

observational study and the survey experiment do not provide direct evidence of the existence of censorship normalization in China. This might require more qualitative evidence or representative longitudinal surveys and will be a task for future research. Another limitation of my current experimental design is that I rely on short-term intense exposure to censorship to measure the treatment effect, which might not resemble the real world. Although previous research in psychology has shown that short-term intense exposure to a stimulus has similar desensitizing effects compared with long-term exposure (Fanti et al. 2009), in reality, citizens are less likely to encounter censorship as intensely as in the experiment and the actual normalization process might take longer than a short survey experiment. Future research should explore more creative designs that resemble the real-world experience of censorship and better capture the long-term normalization process.

# Conclusion

At the beginning of this article, I asked why existing literature that claims censorship awareness will lead to backlash cannot explain the reality that a majority of Chinese citizens are either apathetic toward or supportive of the Chinese government's censorship policy. I pointed out that these existing studies primarily focus on censorship of government criticism and collective action, whereas the targets of censorship might be much broader. Building on the desensitization theory in psychology, I argue that when the range of censorship is expanded beyond politically threatening content to seemingly harmless apolitical topics, citizens are more likely to view censorship as normal and less likely to react as intensely and negatively. In other words, they are more likely to be desensitized to censorship activities when censorship targets both political and non-political content.

The experimental and observational evidence that I presented supports my normalization theory. It shows the possibility that normalization of censorship happens in China and the government attempts to create a benevolent image of censorship policy to downplay the potentially unpopular suppression of political opposition. Moreover, expanding censorship targets has significant effects at the individual level. Citizens exposed to normalized censorship that targets both political and non-political content display significantly less backlash against both the censorship policy and the regime.

The normalization theory can also be applied beyond China. Many authoritarian regimes employ similar strategies of dilution and repetition to justify their censorship apparatus and desensitize their citizens to information control. For example, Esberg (2020) shows how Pinochet's Chile and other right-wing Catholic dictatorships like military-run Argentina and Franco's Spain censored non-political content in addition to political content. By including content the Catholic Church opposed among its censorship targets, Pinochet's regime not only avoided backlash, but instead won support from conservative groups.

Moreover, autocracies are not the only systems of government that engage with censorship. In many democracies including the United States, censorship has become a hotly debated issue in recent years. Future research could extend the theoretical framework here to democracies and explore how regime type affects public reactions to and acceptance of censorship.

Besides censorship, normalization theory also provides an effective explanation of the phenomenon that some repressive policies in authoritarian regimes cause outcries and widespread attention in Western media whereas most people in the authoritarian regime do not have strong reactions to them. For example, Western media have widely reported on China's digital surveillance system powered by millions of digital cameras in almost every corner of major Chinese cities. Similar concerns of surveillance are raised about China's social credit system and mandatory coronavirus tracing app. Yet Chinese citizens do not seem to be bothered much about all the surveillance. On one hand, the primary purpose of these surveillance systems is to prevent crime and ensure public safety, whereas repression of dissent only happens occasionally. This leads most Chinese citizens to believe that these surveillance systems are benevolent. On the other hand, Chinese citizens encounter these surveillance systems every day. CCTV cameras can be seen on almost every street. Repeated exposure has effectively desensitized them, and these surveillance systems that cause outcries in the Western world are just part of the normal life in China.

# References


Anderson, Craig A, Akiko Shibuya, Nobuko Ihori, Edward L Swing, Brad J Bushman, Akira Sakamoto, Hannah R Rothstein and Muniba Saleem. 2010. "Violent Video Game Effects on Aggression, Empathy, and Prosocial Behavior in Eastern and Western Countries: A Meta-Analytic Review." *Psychological Bulletin* 136(2):151.

Arendt, Hannah. 1973. *The Origins of Totalitarianism*. Vol. 244 Houghton Mifflin Harcourt.

Bartholow, Bruce D, Brad J Bushman and Marc A Sestir. 2006. "Chronic Violent Video Game Exposure and Desensitization to Violence: Behavioral and Event-Related Brain Potential Data." *Journal of Experimental Social Psychology* 42(4):532–539.

Boxell, Levi and Zachary Steinert-Threlkeld. 2019. "Taxing Dissent: The Impact of a Social Media Tax in Uganda.".

Cairns, Christopher and Allen Carlson. 2016. "Real-world Islands in a Social Media Sea: Nationalism and Censorship on Weibo during the 2012 Diaoyu/Senkaku Crisis." *The China Quarterly* 225:23–49.

Cantoni, Davide, Yuyu Chen, David Y Yang, Noam Yuchtman and Y Jane Zhang. 2017. "Curriculum and Ideology." *Journal of Political Economy* 125(2):338–392.

Carnagey, Nicholas L, Craig A Anderson and Brad J Bushman. 2007. "The Effect of Video Game Violence on Physiological Desensitization to Real-Life Violence." *Journal of Experimental Social Psychology* 43(3):489–496.

Carpini, Michael X Delli and Scott Keeter. 1996. *What Americans Know About Politics and Why It Matters*. Yale University Press.

Chen, Jidong and Yiqing Xu. 2017. "Why do Authoritarian Regimes Allow Citizens to Voice Opinions Publicly?" *The Journal of Politics* 79(3):792–803.

Chen, Yuyu and David Y Yang. 2019. "The Impact of Media Censorship: 1984 or Brave New World?" *American Economic Review* 109(6):2294–2332.

China Youth Daily. 2015. "Waimei Zhiyi Zhongguo Wangluo Shencha Zhidu Riqu Yange, Lu Wei: Yongci Budang. [China's Internet Censorship Chief Denies Allegations from CNN

Reporter.].” *Sina.com* .

**URL:** *http://news.sina.com.cn/c/2015-12-09/doc-ifxmifzc0923447.shtml*

Dickson, Bruce. 2016. *The Dictator's Dilemma: The Chinese Communist Party's Strategy for Survival*. Oxford University Press.

Efran, Jay S and James E Marcia. 1967. Treatment of Fears by Expectancy Manipulation: An Exploratory Investigation. In *Proceedings of the Annual Convention of the American Psychological Association*. American Psychological Association.

Esberg, Jane. 2020. “Censorship as Reward: Evidence from Pop Culture Censorship in Chile.” *American Political Science Review* 114(3):821–836.

Fanti, Kostas A, Eric Vanman, Christopher C Henrich and Marios N Avraamides. 2009. “Desensitization to Media Violence Over a Short Period of Time.” *Aggressive Behavior: Official Journal of the International Society for Research on Aggression* 35(2):179–187.

Freedom House. 2019. “China Media Bulletin: 2019 Internet Freedom Trends, Shutterstock Censorship, Huawei “Safe Cities” (November 2019).”.

Geddes, Barbara and John Zaller. 1989. “Sources of Popular Support for Authoritarian Regimes.” *American Journal of Political Science* 33(2):319–347.

Goldfried, Marvin R. 1971. “Systematic Desensitization as Training in Self-Control.” *Journal of Consulting and Clinical Psychology* 37(2):228.

Gueorguiev, Dimitar D and Edmund J Malesky. 2019. “Consultation and Selective Censorship in China.” *The Journal of Politics* 81(4):1539–1545.

Guriev, Sergei and Daniel Treisman. 2015. “How Modern Dictators Survive: An Informational Theory of the New Authoritarianism.” *National Bureau of Economic Research* .

Han, Rongbin. 2018. *Contesting Cyberspace in China: Online Expression and Authoritarian Resilience*. Columbia University Press.

Hill, Seth J. 2017. “Learning Together Slowly: Bayesian Learning about Political Facts.” *The Journal of Politics* 79(4):1403–1418.

Hobbs, William R and Margaret E Roberts. 2018. “How Sudden Censorship Can Increase

Access to Information." *American Political Science Review* 112(3):621–636.

Hoerl, Arthur E and Robert W Kennard. 1970. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics* 12(1):55–67.

Huang, Haifeng. 2017. "A War of (Mis)Information: The Political Effects of Rumors and Rumor Rebuttals in an Authoritarian Country." *British Journal of Political Science* 47(2):283–311.

Huang, Haifeng. 2018. "The Pathology of Hard Propaganda." *The Journal of Politics* 80(3):1034–1038.

Huang, Haifeng and Yao-Yuan Yeh. 2019. "Information from Abroad: Foreign Media, Selective Exposure and Political Support in China." *British Journal of Political Science* 49(2):611–636.

Iqbal, Mansoor. 2020. "WeChat Revenue and Usage Statistics (2020)." *Business of Apps* .

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

King, Gary, Jennifer Pan and Margaret E. Roberts. 2014. "Reverse-Engineering Censorship in China: Randomized Experimentation and Participant Observation." *Science* 345:1–10.

Le Cessie, Saskia and Johannes C Van Houwelingen. 1992. "Ridge Estimators in Logistic Regression." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 41(1):191–201.

Lorentzen, Peter. 2014. "China's Strategic Censorship." *American Journal of Political Science* 58(2):402–414.

Marcia, James E, Barry M Rubin and Jay S Efran. 1969. "Systematic Desensitization: Expectancy Change or Counterconditioning?" *Journal of Abnormal Psychology* 74(3):382.

Miller, Blake. 2018. Delegated Dictatorship: Examining the State and Market Forces behind Information Control in China PhD thesis.

Mullinix, Kevin J, Thomas J Leeper, James N Druckman and Jeremy Freese. 2015.

"The Generalizability of Survey Experiments." *Journal of Experimental Political Science* 2(2):109–138.

Nabi, Zubair. 2014. "Resistance Censorship is Futile." *First Monday* .

Ng, Jason Q. 2015. "Politics, Rumors, and Ambiguity: Tracking Censorship on WeChat's Public Accounts Platform." *Munk School of Global Affairs* .

Nie, Ke. 2021. "Disperse and Preserve the Perverse: Computing How Hip-Hop Censorship Changed Popular Music Production in China." *21st Century China Center Research Paper Forthcoming* .

Pan, Jennifer and Alexandra A. Siegel. 2020. "How Saudi Crackdowns Fail to Silence Online Dissent." *American Political Science Review* 114:109–125.

Roberts, Margaret E. 2018. *Censored: Distraction and Diversion Inside China's Great Firewall*. Princeton University Press.

Roberts, Margaret E. 2020. "Resilience to Online Censorship." *Annual Review of Political Science* 23:22.1–22.19.

Shadmehr, Mehdi and Dan Bernhardt. 2015. "State Censorship." *American Economic Journal: Microeconomics* 7(2):280–307.

Svolik, Milan W. 2012. *The Politics of Authoritarian Rule*. Cambridge University Press.

Tai, Yun and King-wa Fu. 2020. "Specificity, Conflict, and Focal Point: A Systematic Investigation into Social Media Censorship in China." *Journal of Communication* 70(6):842–867.

The Paper. 2020. "Weixin Zai Huiying Gonghao Xilie Weiguiwen: Yi Shan 1.59 Wan Pian [WeChat Responded to Censorship: Has Already Censored 15.9K Articles]." *The Paper (Péngpài Newspaper)* .

**URL:** *https://wap.peopleapp.com/article/5398944/5313034?from=timelineisappinstalled=0*

Wang, Dakuo and Gloria Mark. 2015. "Internet Censorship in China: Examining User Awareness and Attitudes." *ACM Transactions on Computer-Human Interaction (TOCHI)* 22(6):1–22.

Wang, Yuhua. 2019. "The Political Legacy of Violence during China's Cultural Revolution."

*British Journal of Political Science* pp. 1–25.

Xinhua. 2018. "Jinghua Wangluo Kongjian, Gongjian Jingshen Jiayuan [Cleaning Online Space; Building Common Homeland]." *Xinhua News Agency* .
**URL:** *http://www.xinhuanet.com/2018-04/18/c*$_1$122704353.*htm*

Young, Lauren E. 2019. "The Psychology of State Repression: Fear and Dissent Decisions in Zimbabwe." *American Political Science Review* 113(1):140–155.