Normalization of Censorship: Evidence from China

Tony Zirui Yang, Emory University and University of Oxford

Authoritarian censorship is traditionally understood as a repressive tool and will cause public backlash against the regime. However, surveys worldwide consistently find that citizens in authoritarian regimes are apathetic toward or even supportive of government censorship. To explain this puzzle, I theorize that citizens are subject to a normalization process. Specifically, individuals become desensitized to censorship activities when the range of censored content expands beyond politically threatening topics, like government criticism and collective action, to other less political issues. Using 28 million censored articles and two original survey experiments in China, I show that (1) a majority of censored articles are indeed unrelated to politically threatening topics, and (2) respondents exposed to the censorship of both political and nonpolitical content report higher support for censorship and the regime than those exposed only to political censorship. These findings highlight how normalizing repressive apparatuses like censorship contributes to authoritarian control and survival.

Government censorship in authoritarian regimes has traditionally been understood as a repressive tool to suppress political oppositions (Gueorguiev and Malesky 2019; Miller 2018; Pop-Eleches and Way 2023), filter unfavorable news (Rozenas and Stukal 2019; Shadmehr and Bernhardt 2015), and hinder collective action (King, Pan, and Roberts 2013, 2014; Lorentzen 2014). As such, the public resists government censorship when they encounter it (Roberts 2018, 2020; Zhu and Fu 2021). For example, when citizens are exposed to censorship events, they express more anger and antiregime sentiment (Pan and Siegel 2020; Roberts 2018, 137), 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 and state media (Gläßel and Paula 2020; Roberts 2018, 144), and even participate in protests against the regime (Boxell and Steinert-Threlkeld 2022). In light of these studies, Roberts (2020) posits that "awareness of censorship [is] essential to resilience to censorship."

However, in China, where the scale of overt censorship activities is by far the largest around the world (Freedom House 2019; Gueorguiev and Malesky 2019; Han 2018; King et al. 2013; Miller 2018), surveys consistently find that Chinese citizens are either apathetic toward or supportive of the regime's censorship apparatus, even when they have experiences with censorship. For example, Dickson (2016, 71) reports that Chinese citizens who have experienced censorship are "rather blasé about it." Wang and Mark (2015) show that a majority (65.6%) of "censorship-aware respondents" are either neutral or supportive of Internet censorship. Studies on other authoritarian regimes such as Russia (Nisbet, Kamenchuk and Dal 2017) and Middle Eastern monarchies (Martin, Martins, and Wood 2016; Wike and Simmons 2015) also find similar phenomena. If awareness of censorship leads to backlash, why do many citizens in authoritarian regimes 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. Although politically threatening content has traditionally been understood as the prime target of censorship (Gueorguiev and Malesky 2019; King et al. 2013, 2014; Miller 2018; Roberts 2018), other seemingly harmless nonpolitical content, such as popular culture (Nie 2021) and pornography (King et al. 2013), is also widely censored in authoritarian regimes. In this study, I argue that when the range of censored content extends beyond highly politically threatening content, such as collective action and government criticism, to other less politically threatening content, citizens are less likely to view censorship as political

Tony Zirui Yang (zirui.yang@nuffield.ox.ac.uk) is an assistant professor of political science at Emory University, Atlanta, GA 30322, and a postdoctoral prize research fellow in politics at Nuffield College, University of Oxford, Oxford, UK OX1 1NF.

Both survey experiments were pre-registered on Open Science Framework. The preanalysis plans can be found at https://osf.io/4htz3 and https://osf.io/4pg8f. Replication files are available in the *JOP* Dataverse (https://dataverse.harvard.edu/dataverse/jop). The empirical analysis has been successfully replicated by the *JOP* replication analyst. An online appendix with supplementary material is available at https://doi.org/10.1086/734239.

Published online June 10, 2025.

  https://doi.org/10.1086/734239

suppression. Rather, censorship becomes viewed as a normal government policy that regulates both political speeches and apolitical content like entertainment, culture, and advertisement. I call such changes in perception *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 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 and less political messages 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 nonpolitical 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, citizens' negative reactions toward censorship, such as anger and antiregime sentiment, should be less likely to occur (Wang and Mark 2015).

To provide evidence for my theory of normalization, I first use observational data to illustrate that non–politically threatening content is being censored on a substantial scale. I collected and categorized more than half a million censored articles from WeChat, China's largest social media platform, as well as more than 27 million censored posts on Weibo, the second-largest platform in China. To ensure the accuracy of my findings, I employed both human coders and deep learning models, specifically the Chinese Bidirectional Encoder Representations from Transformers (BERT) with the Whole Word Masking model (Lu, Pan, and Xu 2021). The results show that collective action, government criticism, and other government-related content only account for approximately 30% of all censored articles. The majority of censored articles are non–politically threatening and include a wide range of nonpolitical topics.

I then conducted two original survey experiments with similar designs in China, in 2020 and 2022 respectively, to test the effect of censoring non–politically threatening content on support for censorship and the regime. In both experiments, I randomly expose respondents to varying amounts of censorship of nonpolitical content. Consistent with the normalization theory, respondents exposed to the censorship of both political and nonpolitical content display significantly less backlash and greater support for the censorship apparatus and the regime, compared with those that are only exposed to the censorship of political content.

This study makes important contributions to the understanding of government censorship in authoritarian regimes. Although I am not the first to argue that authoritarian governments censor nonpolitical content such as pornography (King et al. 2013) and popular culture (Esberg 2020), I challenge the centrality of political censorship in the literature. The results of this study highlight that the censorship of nonpolitical content is crucial to understanding how political censorship can work effectively. My results also bridge 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. Finally, 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; Stukal et al. 2022). Overt censorship, when applied broadly, might desensitize the public to censorship, which subsequently reduces public backlash.

A broader implication of these findings concerns the dynamic of authoritarian control. A wealth of scholarship has investigated how authoritarian governments persuade and threaten people through propaganda, silence dissent through censorship, and prevent uprisings through repression (Arendt 1976; Chen and Xu 2017; Dickson 2016; Guriev and Treisman 2022; Huang 2018; Shadmehr and Bernhardt 2015; Svolik 2012; Young 2019). My study suggests that there might be an additional channel through which authoritarian regimes achieve social control: the normalization of coercive policies. Such a channel is potentially difficult to undermine because citizens are desensitized and subconsciously accept the coercive policy as normal.

NORMALIZATION OF CENSORSHIP: A THEORY

The central argument of this study is that when the range of censorship expands beyond highly politically threatening content, such as collective action and government criticism, to other less politically threatening content, it normalizes the censorship apparatus 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 nonpolitical 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 et al. 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 that 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 on 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 et al. 2007; Marcia, Rubin, and Efran 1969). For example, playing violent video games or watching violent movies 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 et al. 2007). Similarly, if citizens are exposed to censorship in a neutral or positive context, they will view censorship as normal and be 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, can evoke strong reactions, especially when they find an online community of outspoken users who are also angry about the censorship event (Pan and Siegel 2020; Zhu and Fu 2021). In contrast, indirect interaction with censorship, such as observing a web page blocked or discussing censorship online, may yield less negative responses. When users come across deleted posts, they rely on available information, such as the post's title, author, and other users' comments and reactions, to form their beliefs about the censorship event.[1]

If censored content consistently concerns politically threatening topics, users tend to associate censorship events with negative news and government wrongdoings. In contrast, censorship of nonpolitical content dilutes this negative image. When users encounter censorship of nonpolitical content, they update their beliefs and are less likely to associate censorship with hiding politically valuable information or silencing political dissent. Consequently, they are also less likely to witness an angry public reaction against censorship.

Moreover, users frequently engage in discussions about censorship in online conversations, often in neutral and apolitical contexts (Han 2018). For instance, in February 2020, fans of different entertainment stars engaged in heated debates that often involved advocating for the censorship of opposing posts. Such conversations imply that censorship can extend to nonpolitical content related to entertainment stars. Observers of these discussions may be less inclined to view censorship negatively and connect it with political repression.

According to existing surveys, only 9% of citizens have directly experienced censorship (Dickson 2016), even though 69.5% of citizens are aware of censorship (Wang and Mark 2015). Hence, direct interaction with censorship is relatively rare, and most citizens form their beliefs about censorship through indirect interactions. Thus, the range of censored content significantly influences citizens' attitudes toward censorship. When seemingly harmless nonpolitical topics are included among censorship targets, the effectiveness of censorship as a signal for government wrongdoing diminishes. As a result, citizens adjust their beliefs about censored information and are less likely to exhibit significant backlash when encountering censorship. In short, including nonpolitical content among censorship targets dilutes the proportion of politically threatening content and changes citizens' belief that censorship is abnormal and hides politically valuable information such as government wrongdoings. Consequently, negative reactions to censorship of political content are less likely to occur.

Increasing citizens' exposure to censorship

In addition to dilution, censoring nonpolitical content increases the frequency of citizens' exposure to censorship, which

1. The extent of remaining information post-censorship varies among social media platforms. For instance, on WeChat, users can still view the title of a deleted article, but they only discover the content is censored if they attempt to access the full article. Sina Weibo employs a more diverse range of censorship techniques Miller (2018).

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 et al. 2006; Carnagey et al. 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 et al. 2006; Carnagey et al. 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 increased exposures further facilitates dilution. Because citizens regard censorship as normal and common in daily life, they would not deliberately avoid it in conversation as they do with other sensitive topics. The example above about fans of entertainment stars illustrates how censorship could be a normal topic that is frequently mentioned in online conversations. The effect of increased exposure becomes a positive circle that reinforces the belief that censorship is normal.

An important implication of the normalization theory is that normalization is not merely a combination of popular and unpopular events but rather a systematic shift in public perception of censorship, which affects their downstream beliefs of censored content and subsequent behaviors when encountering censorship activities. Once individuals believe censorship is normal, they are less likely to pay attention to censored content and are more likely to dismiss most censored content as not valuable or "deserved to be censored." As such, individuals who believe censorship is normal should show higher support for the censorship apparatus as a whole, not just the nonpolitical part.

Empirical expectations

To summarize the empirical expectations of the theoretical arguments laid out above, I hypothesize that citizens exposed to censorship of both political and nonpolitical content will display greater support for both the censorship apparatus (hypothesis 1) and the regime (hypothesis 2) compared with citizens exposed to censorship of political content only.

Before I test the two main hypotheses, however, I need to demonstrate the possibility that censorship normalization happens in China. In the next section, I combine the existing literature on the Chinese censorship regime with my fieldwork conducted from November 2022 to February 2023 in China to illustrate the institutional environment for the normalization of censorship. I then use data on censored articles to explore one observable implication of the normalization theory: Nonpolitical content accounts for a large proportion of all censored content, that is, Pr(Non-Political Content | All Censored Content) is high. To be clear, this is not to say that nonpolitical content is more likely to be censored than political content, that is, 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 demonstrate that censorship of nonpolitical content happens on a substantial scale.

INSTITUTIONAL DEVELOPMENT OF CHINA'S CENSORSHIP

Although online censorship occurs in almost every authoritarian regime, the range and scale of government censorship in China is by far the largest (Freedom House 2019). To achieve normalization explained in the theory section, the regime often needs a relatively high-capacity censorship apparatus that not only serves the function of political repression but also manages other non–politically threatening content.

Before the 2010s, censorship power in China was fragmented among different authorities (Alshabah 2016; Han 2018; Roberts 2018).[2] During this period, while the Chinese state has been relatively successful in curbing political threats from the Internet to regime stability similar to the Arab Spring, the regulatory fragmentation has led to the inability to exert extensive control over the Internet (Yang 2009). This is reflected in earlier studies showing that Chinese censorship primarily targeted the most dangerous content with collective action potentials while allowing most other online expressions (King et al. 2013, 2014).

In 2014, the Chinese leader Xi Jinping launched an effort to unify and centralize the administrative power of cyberspace, establishing the Central Leading Group for Cybersecurity and Informatization, which is directly chaired by Xi Jinping himself (Tai and Fu 2020). The Leading Group gained control over the Cyberspace Administration of China (CAC), which used to be primarily led by the State Council, and rapidly

2. This period is often referred to as the period where "nine dragons reign together." Censorship power was separated among departments ranging from public security and national security to education, industry, and labor.

expanded its regulatory responsibility, gabbing power from other ministries under the State Council.[3] The centralization of censorship power had resulted in a more aggressive approach to Internet censorship in China. In addition to banning more political content than previously understood (Gueorguiev and Malesky 2019; Miller 2018), many seemingly harmless posts such as tabloid gossip were 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 had traditionally been given freer rein became more systematically censored (Tai and Fu 2020).

Meanwhile, the government's narrative had also become more explicit about its intention to expand the censorship apparatus beyond safeguarding political stability. Before the 2010s, defending national security against foreign subversion had been a central theme of government documents related to Internet censorship, with only a few exceptions such as campaigns against online pornography (Han 2018). This is consistent with the traditional view of censorship as a tool of political repression and a showcase of regime strength. As the censorship power centralized and expanded during the early 2010s, official propaganda rhetoric became more diverse and less politically threatening. Instead, it tended to emphasize the less politically sensitive side of the Internet and highlight the need for extensive Internet regulations.

In the late 2010s and early 2020s, as political and technological competition with the United States and other Western powers intensified, the CAC doubled down on the national security rhetoric, citing foreign subversion threats, while also emphasizing the nonpolitical benefits the authority provided such as digitization and Internet civility.[4] Combining the different pieces of anecdotes above, although censorship activities have been expanding rapidly in recent years, the image of censorship that the Chinese government tries to present has become more benevolent and less politically repressive. The narratives of the central government suggest an association between the expansion and the normalization of censorship.

However, it is worth acknowledging that, even after the centralization of censorship power, the Chinese censorship apparatus is still far from a monolithic system in which every censorship activity stems from the command of a central authority. Instead, it is a comprehensive apparatus that involves many significant decision-makers such as local governments (Lorentzen 2014) and social media companies (Han 2018; Lv and Luo 2018; Miller 2018), which also exert certain levels of influence over what content is censored. The purpose of this study is not to show that the normalization of censorship is a grand strategy designed and implemented by the central government alone. Rather, I aim to highlight the recent expansion of censorship power and increase in censorship activities in China and its downstream normalizing effects on public perception of the censorship apparatus.

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 less politically threatening content occurs on a substantial scale. I collect censored articles from WeChat and Sina Weibo, the two largest social media platforms in China. I then classify the censored articles into nine different topic categories, including three highly political, two moderately political, and four nonpolitical 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 both human coders and natural language processing models, specifically the Chinese BERT with the Whole Word Masking model (Lu et al. 2021).

Data source

My observational study relies on three distinct sources of censorship data from China. Two of these sources are the WeChatScope and WeiboScope datasets, which were collected by a research team at the University of Hong Kong (Tai and Fu 2020; Zhu and Fu 2021). These datasets monitor more than 4,000 WeChat public accounts and more than 118,000 Sina Weibo users, respectively. The third dataset, FreeWeChat, was gathered from the nongovernmental organization GreatFire.org, which monitors more than 34,000 WeChat public accounts in real time. WeChat public accounts are similar to Facebook public pages or Telegram public groups, whereas Sina Weibo is similar to Twitter. Given that WeChat public accounts have a large number of subscribers and posts on Weibo are easily disseminated, both platforms are prime targets of Chinese censorship. Table 1 summarizes the three data sources.

Moreover, both WeChat and Sina Weibo are ideal platforms for analyzing the implication of government censorship on citizens' attitudes because they are the two most popular social media in China. As such, censorship on WeChat

3. According to my conversations with censorship officials during my fieldwork, the then-director of the CAC, Lu Wei, played a critical role in the power-grabbing process of the CAC. His strongman-style leadership helped the CAC become the central coordinator of the entire censorship apparatus.

4. Such parallel messaging is reflected in both official documents published by the CAC and my conversations with central and local censorship officials during my fieldwork.

Table 1. Date Sources for Observational Analysis

	WeChatScope	FreeWeChat	WeiboScope
Platform	WeChat	WeChat	Sina Weibo
Content source	Public Accounts	Public Accounts	Users
Number of censored posts/articles	15,872	533,707	27 million
Data start date	2018-03-01	2016-04-25	2021-05-01
Data end date	2020-05-09	2022-12-26	2022-06-30
Data source	HKU	GreatFire.org	HKU

and Weibo influences a large proportion of the Chinese population, and most Chinese are likely to form their beliefs about censorship via WeChat and Weibo. Using the three data sources provides a hard case to illustrate the existence of large-scale censorship of nonpolitical content. Because all three projects are designed to capture government censorship as conventionally understood, political accounts are overrepresented in the sample (Tai and Fu 2020; Zhu and Fu 2021).[5] Therefore, even if the selected WeChat public accounts and Weibo users might not be representative, the bias is likely to be in the opposite direction of my theoretical expectations.

One caveat to all three data sources is that, like most quantitative censorship data, it only includes post hoc censorship. The articles or posts need to be published on WeChat or Weibo before they can 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" before it can be posted online (King et al. 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 nonpolitical content. Nonetheless, citizens have creative ways to bypass keyword blocking, and post hoc censorship is the most extensive form of censorship (Han 2018; King et al. 2013). As such, I focus on post hoc censorship in this analysis.

Categorization of censored articles

The primary objective of the categorization is to determine the proportion of highly political, moderately political, and nonpolitical content. To define highly political content, I refer to the existing literature on censorship, which identifies three main highly political categories: (1) collective actions, including protests, strikes, rightful resistance, and other collective civil disobedience (King et al. 2013, 2014); (2) government criticism, which entails criticisms of the central or local government, as well as state-owned enterprises (Gueorguiev and Malesky 2019; Lu et al. 2021; Tai and Fu 2020); and (3) other government-related content, such as discussion of political leaders and political rumors (Miller 2018), or progovernment content that nonetheless mentions sensitive political topics or figures (King et al. 2013, 2014, 2017; Lu et al. 2021; Stukal et al. 2022).

In addition to the highly political categories, there are two moderately political categories that may still be sensitive to the government: business and foreign events. The business category includes articles discussing private companies, whereas the foreign events category covers content related to foreign countries, provided that it does not directly reference China or the Chinese government's economic and foreign policies. Examples of business content include investment tips in the stock market, and examples of foreign events include discussions of the domestic politics of foreign countries. Notably, none of the existing censorship studies consider these categories as politically threatening to the Chinese government (Gueorguiev and Malesky 2019; King et al. 2013; Miller 2018; Tai and Fu 2020). However, some studies on authoritarian propaganda imply that business and foreign news might be relevant for autocrats' legitimacy and political survival (Mattingly and Yao 2022; Rozenas and Stukal 2019). As such, I first follow the censorship literature and only treat collective action and government-related content as political, and then include the two additional categories to provide a more conservative measure of nonpolitical content.

Lastly, I track four nonpolitical categories that are unlikely to pose a threat to the Chinese government: entertainment, advertisement, culture, and other content. Examples of these nonpolitical categories include tabloid gossip, product promotions, and stories of cultural figures. To ensure consistency and accuracy, I adopt the coding rubrics primarily from Miller

5. The WeChatScope and WeiboScope projects primarily include 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 the Baidu search engine (Tai and Fu 2020; Zhu and Fu 2021).

(2018), which provides a comprehensive and up-to-date categorization of censored content in China. Notably, I make one key modification: The nine categories are mutually exclusive. This simplifies both the categorization process and the interpretation of results.[6]

To categorize the vast number of censored articles, I first employ human coders to annotate a training set. I randomly sampled 2,500 articles from WeChatScope, and 5,000 articles from the remaining two data sources, stratified by the creation year. Two native Chinese coders have coded the training set independently. The Cohen's κ between the two coders is 0.80, higher than the commonly applied criteria of 0.70 for intercoder reliability tests. In cases where the two coders disagreed, the author acted as an arbitrator to settle the dispute.

To select the best classification model, I utilized the training data to evaluate the performance of nine different machine-learning models through out-of-sample cross-validation.[7] The fine-tuned pretrained Chinese BERT model is by far the best-performing model, achieving an in-sample accuracy of 0.96 and an out-sample balanced macro F1 score of 0.72. Finally, I use the fine-tuned Chinese BERT model to predict the categories of censored content from all three sources. For more details about the performance of BERT and alternative models, see appendix E.

Results

Table 2 reports the predicted proportion of censored articles and posts classified by topic category across the three different data sources. Consistent with the empirical expectations, a substantial proportion of censored articles are unrelated to politically threatening topics. Let me first focus on the highly political categories highlighted by the existing literature. As illustrated in the top section of table 2, only about 30% of all censored articles across all three data sources pertain to collective action and government-related topics. In contrast, the remaining content that is less politically threatening constitutes the vast majority of censored articles on both WeChat and Sina Weibo. It is important to note that censorship of collective action and government criticism, which are the two categories that have the highest likelihood of generating popular backlash against the regime (Pan and Siegel 2020; Roberts 2020), only occurs about 20% of the time when citizens experience censorship events. In other words, it is much more common for Chinese citizens to encounter censorship of nonpolitical content than censorship of government criticism and collective action. This would significantly lower their expectations that censorship events primarily target political oppositions and are useful for uncovering government wrongdoings. As a consequence, citizens would be more likely to believe that censorship is normal.

Even when considering a broader definition of political content that includes moderately political categories such as business and foreign events, which, as previously explained, explicitly exclude content related to the Chinese government, political content still constitutes only approximately half of all censored content. On Weibo, censored political content is particularly scarce, with only one-third of the censored content in WeiboScope being categorized as highly or moderately political. Additionally, the censorship of nonpolitical content is not concentrated on a single category. For instance, the WeChatScope data recorded around 16.87% of entertainment content, 10.69% of advertisements, and 12.90% of cultural, local, and religious content. The other two censorship datasets also present similar patterns. These findings suggest that the censorship of nonpolitical content is comprehensive and broad, rather than being limited to a specific type of nonpolitical content, such as popular culture (Esberg 2020) or pornography (King et al. 2013), indicating that its objectives extend beyond merely content moderating inappropriate content to a more systematic control over less politically sensitive areas.[8]

The observation that a considerable percentage of censored articles are less politically threatening in nature is consistently observed over time. Figure 1 presents the time series data depicting the proportion of collective action, government-related articles, and other less politically threatening topic categories. Even during the initial stages of the COVID-19 outbreak (February 2020) and later, during COVID policy U-turns (November 2022), when government suppression of online criticism was at its peak, non–politically threatening content still constituted more than half of all censored articles.

To validate the text analysis results, I randomly sampled 1,000 machine-categorized nonpolitical articles and hired a different undergraduate research assistant from China to identify whether there were any hidden or explicit political messages in them. The results of this validation exercise suggest that 91.2% of the machine-categorized nonpolitical content is indeed irrelevant to politically threatening topics. Even after correcting for the 10% miscategorization, a majority of the censored content will still be non–politically threatening.

6. For the detailed explanation of each topic category and coding process, see appendix D.

7. These models include a fine-tuned pretrained Chinese BERT with the Whole Word Masking model, a logistic regression model with a ridge estimator, a pattern learning and matching (PaLM) model, an ensemble classifier model, a random forest model, a decision tree model, an extreme gradient boosting (XGBoost) model, a neural network model, and a word embedding model using Word2vec.

8. Appendix D uses structural topic models to provide further details regarding the content in each specific topic category.

Table 2. Predicted Proportion of Censored Articles/Posts by Topic Category

General Category	Specific Category	WeChatScope	FreeWeChat	WeiboScope
Highly political	Collective action	1.78%	.78%	.52%
	Govt criticism	18.23%	11.79%	11.06%
	Other govt-related	11.84%	14.03%	13.41%
	Total	31.85%	26.60%	24.99%
Moderately political	Business	13.87%	9.95%	6.87%
	Foreign	5.89%	6.96%	5.12%
	Total	19.76%	16.91%	11.99%
Nonpolitical	Entertainment	16.87%	19.34%	21.71%
	Advertisement	10.69%	10.90%	16.32%
	Culture	12.90%	20.63%	17.72%
	Others	7.92%	5.61%	7.28%
	Total	48.38%	56.48%	63.03%

Note. WeChatScope data contains 15,872 censored articles. FreeWeChat data contains 533,707 censored articles. WeiboScope data contains approximately 27 million censored posts.

This also alleviates the concern of political euphemism or other strategies citizens use to circumvent censorship, which could potentially mislead the text analysis results (Han 2018; King et al. 2013).

The implications of the empirical results are twofold. First, the results from both WeChat and Weibo 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 nonpolitical content

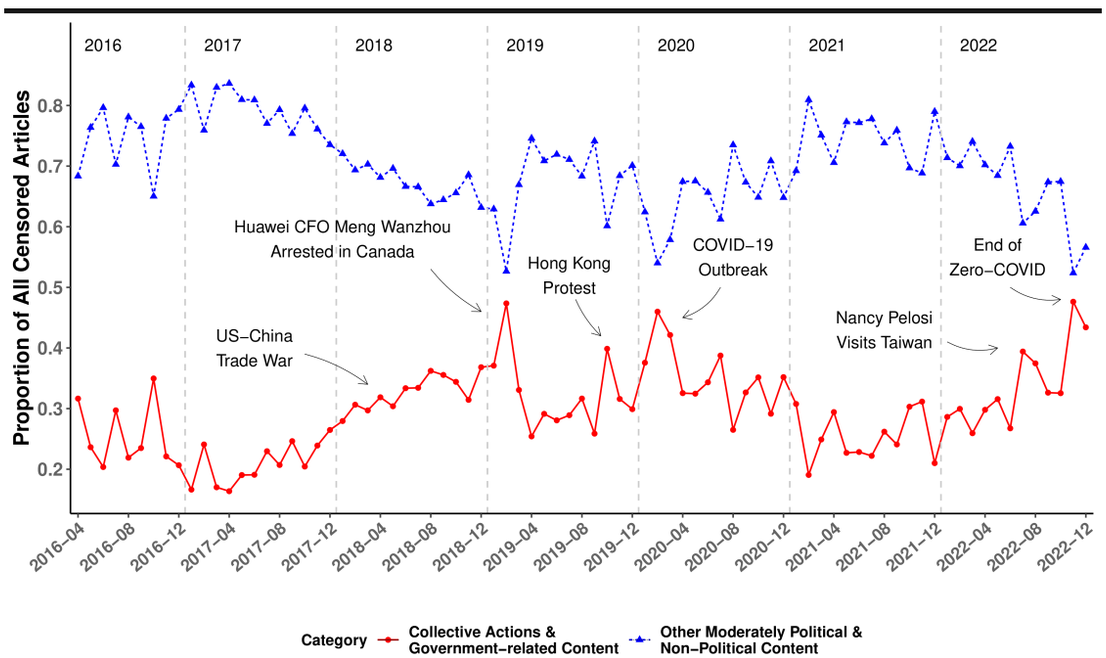

Figure 1. Time series of all censored articles/posts by topic category. All three data sources are weighted equally. Other moderately political and nonpolitical content includes business, foreign events, entertainment, advertisement, cultures, and others.

implies that the censorship of nonpolitical content is broad and comprehensive rather than narrow and focused on one specific type of nonpolitical content such as popular culture (Esberg 2020) or pornography (King et al. 2013). This keeps the red line fuzzy and gives the government leeway to include politically threatening content in broadly targeted censorship activities without drawing too much attention to the political content (Han 2018).

Second, the results imply that it is common for Chinese citizens to encounter censorship of nonpolitical content. As a result, their expectations of censorship outcomes are likely to be shaped by these experiences of censorship of nonpolitical content, which provides favorable conditions for the desensitization process to take place.

NORMALIZING EFFECTS OF CENSORING NONPOLITICAL CONTENT: SURVEY EXPERIMENTS

The previous section illustrates the possibility that censorship normalization happens by showing the existence of large-scale censorship of less politically threatening content. To test the two main hypotheses, I conducted two original survey experiments in China with similar experimental designs. In both experiments, I manipulated the topics of censorship targets and the frequency of censorship activities that participants were exposed to. I then use their ex post evaluations of the censorship apparatus and the Chinese regime to measure the normalizing effects of censorship.

Participants

The first survey experiment was conducted in December 2020 with 612 respondents and the second was conducted in December 2022 with 3,314 respondents.[9] The participants in both experiments were recruited from a Chinese 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 is a reliable way to recruit participants for survey experiments (Mullinix et al. 2015) and Chinese online platforms might even be better 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).[10] To further ensure sample quality, I used attention checks to screen the respondents at the beginning of the surveys. Both samples cover a wide range of socioeconomic backgrounds and are similar to the Chinese demographic in terms of gender, age, and regional distributions. However, like many other online surveys in China, the samples are richer and better educated.

Experimental design

The purpose of the experiments is to compare the downstream effects of censorship that primarily targets politically threatening content with censorship that targets both political and nonpolitical content. Both experiments have similar designs and include three components. First, participants answered pretreatment questions about their socioeconomic backgrounds and political predispositions. Second, participants were randomly assigned to either a control group, where they were exposed to only censorship of political content, or a treatment group, where they were exposed to censorship of both political and nonpolitical content.[11] Finally, respondents answered posttreatment questions about their attitudes toward the censorship apparatus and the regime.

To expose participants to censorship, I asked respondents to read 10 snippets of WeChat articles, presented one at a time with only the title and the 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, and no further information was provided to the respondents. Figure 2 shows one of the snippets with and without a censorship label.[13] Among the 10 snippets, 6 were about nonpolitical topics, whereas 4 were about government criticism and collective action. The order of the snippets was randomized.

I validated the appropriateness of my choice of article snippets in two ways. First, I consulted a panel of China scholars. None of them thought that any of the snippets were absurd or 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 inappropriate, then this would be reflected in an unusually high or low level of interest. None of the snippets stands out as exceptionally high

9. Both experiments were approved by the Institutional Review Board (IRB ID: 202011148) at Washington University in St. Louis and were preregistered on the Open Science Framework.

10. At the end of the survey, respondents were allowed to leave comments on the survey questions. The comments suggest that participants were excited rather than scared by political surveys.

11. In the second experiment, there was an additional blank control group where respondents were not shown any censorship at all.

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. The details of the article snippets can be found in appendix C.

13. In reality, if an article is censored by WeChat, users can only see the title. 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 snippet with and without censorship label in the experiment. The label reads "This article was blocked by WeChat due to violations of Internet law."

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 nonpolitical 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 nonpolitical snippets were also 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 nonpolitical 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 increased exposure laid out in the theory. Table 3 summarizes when and where the censorship label occurred in the treatment and control groups. 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).

Measurement

In both survey experiments, to measure support for censorship apparatus, I asked the respondents whether they agree or disagree that "the government should actively control the Internet and remove content that it deems inappropriate." This is a direct test of hypothesis 1 and the wording is an adaptation inspired by similar questions in the limited number of existing studies on public attitude toward censorship in China (Dickson 2016; Roberts 2018; Wang and Mark 2015). I avoided directly using the word "censorship" because it might be offputting to respondents. To measure regime support, I directly borrowed questions measuring assessment of the government, overall satisfaction, and willingness to protest from Huang (2018). For the assessment of the government, I

Table 3. Explanation of Treatment

Snippet #	Topic Category	Control Group	Treatment Group
1	Political	Censorship label	Censorship label
2	Political	Censorship label	Censorship label
3	Political	Censorship label	Censorship label
4	Political		
5	Nonpolitical		Censorship label
6	Nonpolitical		Censorship label
7	Nonpolitical		Censorship label
8	Nonpolitical		
9	Nonpolitical		
10	Nonpolitical		

Note. The order of the snippets was randomized. The treatment group was exposed to additional censorship of nonpolitical snippets.

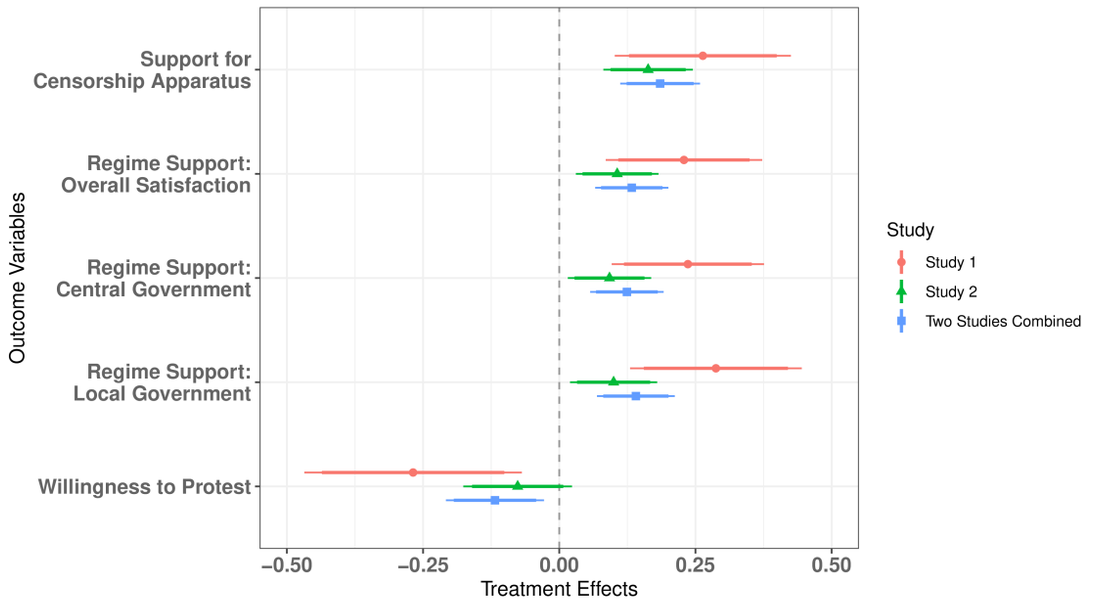

Figure 3. Treatment effects of additional censorship of nonpolitical content on support for the censorship apparatus and the regime (difference-in-means). All outcome variables were measured on a five-point scale. The dots are the difference in means between the treatment and control groups. Bars indicate the corresponding 90% and 95% confidence intervals. *P* values of all coefficients, except the willingness to protest in study 2, are still significant at the conventional level after the Benjamini–Hochberg correction.

asked separately about the central government and local government. Because discussing local government is less sensitive than the 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. All outcome variables were measured on a five-point scale.

I used eight pretreatment covariates to check the balance between the treatment and control groups. They are also included in the regression analyses. Among the covariates, four are demographic variables, including age group, income, gender, and education. The remaining four covariates are ideology, party membership, political interests, and social media usage, which measure participants' political predispositions and Internet usage.

Results

Figure 3 reports the difference in means between the treatment groups and the control groups on individual support for the censorship apparatus and the regime in each of the two experiments as well as for the two studies combined.[14] As shown in figure 3, the results are consistent across the two experiments. In particular, respondents exposed to censorship of both political and nonpolitical content are reliably more supportive of the censorship apparatus (study 1: $\beta = 0.264$, $se = 0.082$; study 2: $\beta = 0.163$, $se = 0.042$; combined: $\beta = 0.185$, $se = 0.037$). Such results are robust to multiple hypotheses testing correction using the Benjamini-Hochberg method (see app. B) and indicate that the range of censorship targets matters for public reactions toward censorship. Additional exposure to censorship of nonpolitical content reduces respondents' backlash against the censorship apparatus even if they are also exposed to censorship of political content.

To further demonstrate the magnitude of the treatment effects, I translate the raw coefficients into proportions of respondents supporting the censorship apparatus. As shown in figure 4*A*, combining the two studies, approximately 58.4% of the respondents in the treatment groups are strongly or somewhat supportive of the censorship apparatus, compared with only 51.3% of those in the control groups. Additional censorship of nonpolitical content increases support by 7.1 percentage points.

For attitudes toward the regime, combining both studies, respondents exposed to censorship of both political and nonpolitical content express higher overall satisfaction with the regime ($\beta = 0.133$, $se = 0.034$), higher support for both central ($\beta = 0.124, se = 0.034$) and local governments ($\beta = 0.141, se = 0.036$), and a lower willingness to protest ($\beta = -0.118, se = 0.046$). The effect sizes are smaller than

14. Additional regression results with pretreatment covariates are reported in appendix B. The results are consistent with the main results presented in the main article.

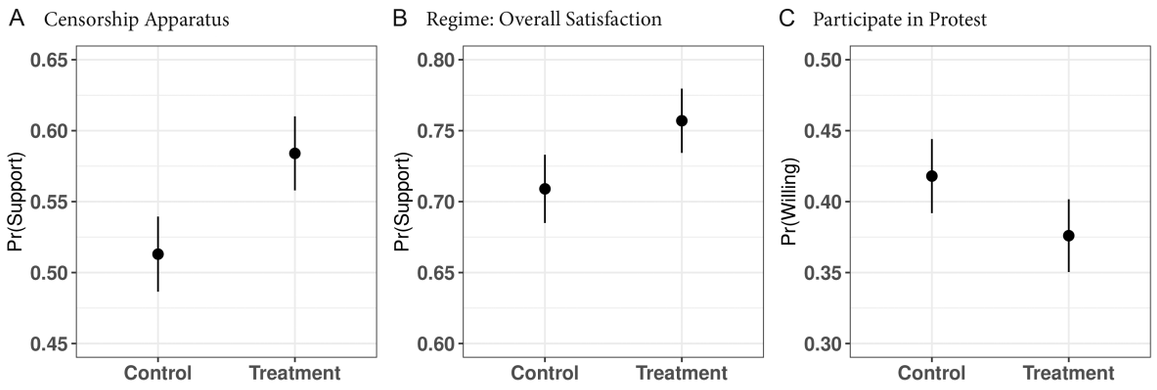

Figure 4. Substantive shifts in support for the censorship apparatus and the regime. The dots are the proportion of respondents who somewhat or strongly support censorship or the regime and are somewhat or definitely willing to participate in protests (two studies combined). The bars refer to the corresponding 95% confidence intervals.

the effects on support for censorship but still considerable given the fact that baseline support for the regime is already high in the control group. As shown in figure 4*B*, approximately 71% of the respondents are already supportive of the regime in the control group. As such, a 5-percentage-point increase in regime support (76% among the treated) and a 5-percentage-point decrease in willingness to protest (from 42% to 37%) are substantial in magnitude.

Discussion

An important issue I would like to address is whether the positive results are artifacts of preference falsification or social desirability bias (Kuran 1997), induced by the heightened censorship in the treatment group. To address this concern, study 2 incorporates an additional blank control group, where respondents are not exposed to censorship at all.[15] If the results in the main analyses are indeed driven by preference falsification induced by more censorship, we should expect the blank control group, without any censorship and therefore no priming of social desirability bias, to exhibit the lowest level of support. Moreover, we should expect censorship of political content to induce more severe preference falsification than nonpolitical content.

However, as illustrated in figure 5, these expectations do not hold. Compared with the blank control group without censorship (left bars in each panel), respondents exposed to censorship of political content (middle bars) reported lower support for the regime and its censorship apparatus, as well as a higher willingness to protest. These findings are consistent with existing studies that highlight the backlash effects stemming from political censorship (Pan and Siegel 2020; Roberts 2018), indicating that respondents are not coerced by the censorship primes into downplaying their aversion to censorship.

In contrast, when respondents are exposed to additional censorship of nonpolitical content (right bars in each panel of fig. 5), their levels of support are indistinguishable from the blank control group without censorship (left bars). Such results demonstrate how additional censorship of nonpolitical content normalizes public perception of the censorship apparatus as if the "new normal" is equivalent to the initial environment without censorship.

To further alleviate the concern that the treatment group might induce additional preference falsification compared with the control group, study 2 also uses a list experiment to measure implicit support for censorship. The results show that overall, at most 7% of the respondents falsified their preference, and the treatment group has roughly the same level of preference falsification as the control group (see app. B). Nevertheless, it is worth noting that fully isolating the impact of preference falsification might be difficult, and readers should still interpret these results with caution.

The experimental results have several implications. First, the results confirm that expanding the range of censorship targets beyond politically threatening content to seemingly harmless nonpolitical content significantly increases public support for both the censorship apparatus and the regime. Such an expansion seems to be a plausible explanation of why most Chinese citizens are apathetic toward or supportive of government censorship. When exposed to censorship of nonpolitical content, citizens update their beliefs about censorship and are more likely to view censorship as normal rather than repressive.

Second, the results show that repressive apparatuses like censorship might be popular and can increase regime support.

15. Respondents in the blank control group still read the same 10 social media snippets.

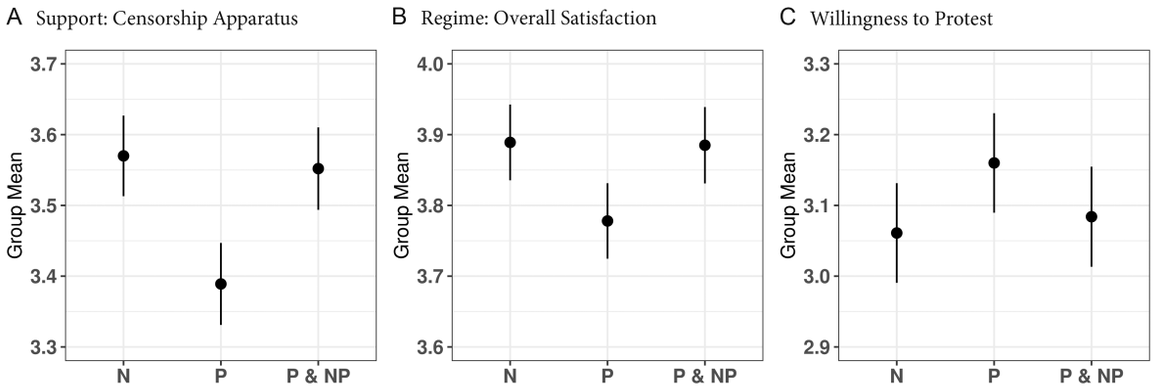

Figure 5. Political censorship decreases support for censorship and the regime and increases willingness to protest, whereas additional censorship of nonpolitical content brings support back to initial levels without censorship. All outcome variables were measured on a five-point scale. The dots are the group means. Bars indicate the corresponding 95% confidence intervals. N, no censorship (blank control group); P, censorship of political content only (control group); P & NP: censorship of both political & nonpolitical content (treatment group).

This suggests that normalization is an independent channel of authoritarian control, different from persuasion, which highlights the achievements of the government, and repression, which showcases regime strength and deters dissent.

ALTERNATIVE EXPLANATIONS AND LIMITATIONS

Before concluding, let me also discuss an important alternative explanation. Although the main experimental results demonstrate that censorship of nonpolitical content leads to higher support, it might not be due to the normalization of censorship. Instead, individuals might simply favor censorship of nonpolitical content, particularly when they believe the nonpolitical content deserves to be censored. This is similar to public support for content moderation in democratic settings, such as removing misinformation, vulgar language, or pornography. Hence, bundling the censorship of political and nonpolitical content increases the overall support for the censorship apparatus.

Although it is both possible and probable that citizens might like censoring some types of nonpolitical content, it does not fully explain the phenomenon. As shown in the observational analysis, the range of nonpolitical content is much broader than merely inappropriate content. In the experiment, I also validated the appropriateness of the nonpolitical snippets and excluded those that were overtly absurd or inappropriate. More importantly, individuals who believe the censorship apparatus is normal should show higher support for the censorship of both political content and nonpolitical content, whereas individuals who believe it to be political repression should show lower support for both. In contrast, the alternative bundling explanation predicts that support for censorship of political content remains constantly low and support for censorship of nonpolitical content remains constantly high.

To obtain a better sense of whether the normalization theory explains the treatment effects above, in study 1, I asked respondents an additional question about whether they think "government control of the Internet is normal." I find individuals in the treatment group are significantly more likely to believe that censorship is normal ($\beta = 0.243$, $p = 0.002$), and this belief is a significant mediator between the treatment and support for censorship (ACME $= 0.149$, $p = 0.002$), directly supporting my normalization theory.

To further distinguish the normalization theory from the bundling argument, in study 2, I included two additional outcomes measuring support for the censorship of political content and nonpolitical content respectively. First, comparing the blank control and control groups, I find that respondents in the control group with exposure to political censorship express significantly lower support for the censorship of not only political content ($\beta = -0.126$, $p = 0.004$) but also nonpolitical content ($\beta = -0.095$, $p = 0.035$). Then, comparing the control and treatment groups, I find respondents in the treatment group express significantly higher support for the censorship of nonpolitical content ($\beta = 0.101$, $p = 0.026$). However, albeit in the correct direction, the effect on political content ($\beta = 0.064$, $p = 0.155$) is not statistically significant at the conventional level. Moreover, I asked whether they think censorship targets political content, a proxy for the perceived level of repressiveness, and also find it to be a significant mediator both between the blank control and the control (ACME $= -0.017$, $p = 0.004$) and between the control and the treatment at the 0.1 level (ACME $= 0.011$, $p = 0.084$).

In summary, these findings demonstrate that support for nonpolitical censorship is not consistently high, and support for political censorship is not consistently low either. Instead, it appears that specific measures of support are shaped by individuals' overarching perceptions of the censorship apparatus.

Indeed, causal mediation analyses reveal that these overarching perceptions mediate the treatment effects. Nevertheless, the results do not completely rule out the bundling argument, and it is possible that both explanations are not mutually exclusive. In essence, bundling effects could be occurring alongside normalization.

Apart from alternative explanations, there are also a few potential limitations to both the normalization theory and the empirical analyses. First, theoretically, does censoring *any* nonpolitical content have the normalizing effect? It is well established in the political science literature that people care more about nonpolitical issues like sports and entertainment than political issues (Delli Carpini and Keeter 1996). Therefore, it is logical to argue that censoring some of the popular nonpolitical topics might lead to backlash as well. Consistent with this logic, Hobbs and Roberts (2018) shows that the censorship of nonpolitical content at a large scale, such as blocking certain platforms completely, is not without risk of public backlash. To avoid backlash, there may still be strategic considerations in choosing which nonpolitical topics to censor. This remains an important topic for future research to explore.

It is worth noting that backlash against censorship can still occur, especially during critical moments marked by sudden increases in political suppression. For example, during the initial COVID outbreak in 2020 and the later COVID policy U-turn in 2022, the extensive suppression of political information did trigger public outrage, even in the presence of nonpolitical censorship. This indicates that times of crisis can potentially undermine the effectiveness of normalization, precisely at the moments when the regime relies on it most.

Finally, my study relies on short-term, intense exposure to censorship to detect the normalizing effect, which might not resemble the real world. Although previous research in psychology has shown that short-term, intense exposure 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. The fact that I found consistent results in both survey experiments two years apart, the first in 2020 and the second in 2022, alleviates part of the concerns. However, it might still be better to use longitudinal survey data to identify long-term normalization.

CONCLUSION

At the beginning of this study, I asked why existing literature that claims censorship awareness will lead to backlash cannot explain the reality that most 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 are 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 intensely and negatively.

The experimental and observational evidence that I presented supports my normalization theory. It shows the possibility that the normalization of censorship happens in China and the government attempts to create a benevolent image of the censorship apparatus. Moreover, expanding censorship targets has significant effects at the individual level. Citizens exposed to normalized censorship that targets both political and nonpolitical content display significantly higher support for both the censorship apparatus and the regime.

The normalization theory can also be applied beyond China. Many authoritarian regimes employ similar strategies of diluting the repressive image of the censorship apparatus and desensitizing their citizens to information control. Moreover, autocracies are not the only systems of government that engage with censorship. 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 as well as China's social credit system (Kostka 2019). 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. Repeated exposure has effectively desensitized them, and these surveillance systems that cause outcries in the Western world are just part of normal life in China.

ACKNOWLEDGMENTS

I thank Margit Tavits, Deniz Aksoy, Taylor Carlson, Ted Enamorado, Haifeng Huang, Carly Wayne, Jane Esberg, Mai Hassan, Jen Pan, Alex Siegel, Molly Roberts, Afiq bin Oslan, Ben S. Noble, Jeremy Siow, İpek Ece Şener, Taylor Damann, Rex Weiye Deng, Dahjin Kim, and seminar participants at NYU, Harvard, and Indiana for their comments and suggestions.

King-wa Fu, Martin Johnson, Haohan Chen, Dixin Li, Guanzhen Su, and Paco Yip provided generous help in the data collection process. I am grateful for the financial support from the Centennial Center at the American Political Science Association, the Weidenbaum Center at Washington University, the Chiang Ching-Kuo Foundation, and the Institute for Humane Studies.

REFERENCES

Alshabah, Nabli. 2016. "Information Control 2.0: The Cyberspace Administration of China Tames the Internet." *Merics China Monitor*. https://merics.org/en/report/information-control-20

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. 1976. *The Origins of Totalitarianism*. Harcourt Brace Jovanovich.

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): 532–39.

Boxell, Levi, and Zachary Steinert-Threlkeld. 2022. "Taxing Dissent: The Impact of a Social Media Tax in Uganda." *World Development* 158: 105950.

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.

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): 489–96.

Chen, Jidong 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): 2294–332.

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

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

Esberg, Jane. 2020. "Censorship as Reward: Evidence from Pop Culture Censorship in Chile." *American Political Science Review* 114 (3): 821–36.

Fanti, Kostas A., Eric Vanman, Christopher C. Henrich, 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): 179–87.

Freedom House. 2019. "China Media Bulletin: 2019 Internet Freedom Trends, Shutterstock Censorship, Huawei 'Safe Cities' (November 2019)." https://freedomhouse.org/report/china-media-bulletin/2020/china-media-bulletin-2019-internet-freedom-trends-shutterstock.

Gläßel, Christian, and Katrin Paula. 2020. "Sometimes Less Is More: Censorship, News Falsification, and Disapproval in 1989 East Germany." *American Journal of Political Science* 64 (3): 682–98.

Gueorguiev, Dimitar D., and Edmund J. Malesky. 2019. "Consultation and Selective Censorship in China." *The Journal of Politics* 81 (4): 1539–45.

Guriev, Sergei, and Daniel Treisman. 2022. *Spin Dictators: The Changing Face of Tyranny in the 21st Century*. Princeton University Press.

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

Hobbs, William R., and Margaret E. Roberts. 2018. "How Sudden Censorship Can Increase Access to Information." *American Political Science Review* 112 (3): 621–36.

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): 1034–38.

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): 611–36.

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* 107:326–43.

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2014. "Reverse-Engineering Censorship in China: Randomized Experimentation and Participant Observation." *Science* 345 (6199): 1251722.

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2017. "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument." *American Political Science Review* 111 (3): 484–501.

Kostka, Genia. 2019. "China's Social Credit Systems and Public Opinion: Explaining High Levels of Approval." *New Media & Society* 21 (7): 1565–93.

Kuran, Timur. 1997. *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Harvard University Press.

Lorentzen, Peter. 2014. "China's Strategic Censorship." *American Journal of Political Science* 58 (2): 402–14.

Lu, Yingdan, Jennifer Pan, and Yiqing Xu. 2021. "Public Sentiment on Chinese Social Media During the Emergence of COVID-19." 21st Century China Center Research Paper #2021-4.

Lv, Aofei, and Ting Luo. 2018. "Asymmetrical Power Between Internet Giants and Users in China." *International Journal of Communication* 12:3877–95.

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

Martin, Justin D., Ralph J. Martins, and Robb Wood. 2016. "Desire for Cultural Preservation as a Predictor of Support for Entertainment Media Censorship in Saudi Arabia, Qatar, and the United Arab Emirates." *International Journal of Communication* 10:23.

Mattingly, Daniel C., and Elaine Yao. 2022. "How Soft Propaganda Persuades." *Comparative Political Studies* 55 (9): 1569–94.

Miller, Blake. 2018. "Delegated Dictatorship: Examining the State and Market Forces behind Information Control in China." PhD diss., University of Michigan, Ann Arbor.

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): 109–38.

Nabi, Zubair. 2014. "Resistance Censorship is Futile." *First Monday*. https://firstmonday.org/ojs/index.php/fm/article/view/5525.

Ng, Jason Q. 2015. "Politics, Rumors, and Ambiguity: Tracking Censorship on WeChat's Public Accounts Platform." Munk School of Global Affairs. https://citizenlab.ca/2015/07/tracking-censorship-on-wechat-public-accounts-platform/.

Nie, Ke. 2021. "Disperse and Preserve the Perverse: Computing How Hip-Hop Censorship Changed Popular Music Genres in China." *Poetics* 88:101590.

Nisbet, Erik C., Olga Kamenchuk, Aysenur Dal. 2017. "A Psychological Firewall? Risk Perceptions and Public Support for Online Censorship in Russia." *Social Science Quarterly* 98 (3): 958–75.

Pan, Jennifer, and Alexandra A. Siegel. 2020. "How Saudi Crackdowns Fail to Silence Online Dissent." *American Political Science Review* 114:109–25.

Pop-Eleches, Grigore, and Lucan A. Way. 2023. "Censorship and the Impact of Repression on Dissent." *American Journal of Political Science* 67 (2): 456–71.

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 (1): 401–19.

Rozenas, Arturas, and Denis Stukal. 2019. "How Autocrats Manipulate Economic News: Evidence from Russia's State-Controlled Television." *The Journal of Politics* 81 (3): 982–96.

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

Stukal, Denis, Sergey Sanovich, Richard Bonneau, and Joshua A. Tucker. 2022. "Why Botter: How Pro-Government Bots Fight Opposition in Russia." *American Political Science Review* 116 (3): 843–57.

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

Tai, Yun, King-wa Fu. 2020. "Specificity, Conflict, and Focal Point: A Systematic Investigation into Social Media Censorship in China." *Journal of Communication* 70 (6): 842–67.

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.

Wike, Richard, and Katie Simmons. 2015. "Global Support for Principle of Free Expression, But Opposition to Some Forms of Speech." Pew Research Center. https://www.pewresearch.org/global/2015/11/18/global-support-for-principle-of-free-expression-but-opposition-to-some-forms-of-speech/.

Yang, Guobin. 2009. *The Power of the Internet in China: Citizen Activism Online*. Columbia University Press.

Young, Lauren E. 2019. "The Psychology of State Repression: Fear and Dissent Decisions in Zimbabwe." *American Political Science Review* 113 (1): 140–55.

Zhu, Yuner, and King-wa Fu. 2021. "Speaking Up or Staying Silent? Examining the Influences of Censorship and Behavioral Contagion on Opinion (Non-)Expression in China." *New Media & Society* 23 (12): 3634–55.