Original Article

Comparative Political Studies
2024, Vol. 0(0) 1–27

DOI: 10.1177/00104140241302762
journals.sagepub.com/home/cps
Sage

Participatory Censorship in Authoritarian Regimes

Tony Zirui Yang[1,2]

Abstract

Contrary to the conventional top-down view of government censorship, this study argues that ordinary citizens in authoritarian regimes frequently participate in censorship by reporting online content. I hypothesize that such participation in censorship partially explains the high level of public support for censorship found in existing surveys. Using an original survey in China, I demonstrate that participation in censorship is indeed prevalent, with over half of the respondents self-report having previously flagged online content, and that such participation is positively correlated with support for the censorship apparatus. To causally test the hypothesis, I conduct a pre-registered experiment using custom-engineered, simulated social media pages to manipulate reporting behavior. The results show that respondents encouraged to report simulated posts display significantly higher support for the censorship apparatus. This study highlights the role of ordinary citizens in facilitating authoritarian control and explains why repressive apparatus like censorship can be popular with the population.

Keywords

censorship, authoritarian regime, China, public participation, public opinion

[1]Emory University, Atlanta, GA, USA
[2]University of Oxford, Oxford, UK

Corresponding Author:
Tony Zirui Yang, Department of Political Science, Emory University, Atlanta, GA 30322, USA; Nuffield College, University of Oxford, Oxford OX1 1NF, UK.
Email: zirui.yang@nuffield.ox.ac.uk

Data Availability Statement included at the end of the article

Introduction

The conventional wisdom on government censorship in authoritarian regimes emphasizes its top-down nature, regarding it as a tool used by autocrats to silence criticism and prevent collective action that could destabilize the regime (Gueorguiev & Malesky, 2019; King et al., 2013; Miller, 2018; Pop-Eleches & Way, 2021; Roberts, 2018; Shadmehr & Bernhardt, 2015). Given this understanding, existing research has extensively studied the ways in which ordinary citizens circumvent and resist government censorship activities (Chang et al. 2022; Chen & Yang, 2019; Gläßel & Paula, 2020; Han, 2018; Hobbs & Roberts, 2018; Pan & Siegel, 2020; Roberts, 2018, 2020).[1]

However, contrary to conventional views, citizens in authoritarian regimes frequently participate in censorship by reporting online content. For example, in Turkey, Twitter users systematically weaponize the report function against political opponents (Tufekci, 2017). Similarly, in Russia, hundreds of users maliciously reported supporters of Ukraine on Facebook just before the Russian invasion of Ukraine (Nimmo & Agranovich, 2022). In China, social media platforms explicitly ask their users to report each other for "violating Internet laws" (Jiang, 2021) and reward those who report with "credit scores" (Cook, 2019). Furthermore, surveys across the world demonstrate significant levels of popular support for censorship, especially in authoritarian regimes like China (Dickson, 2016; Wang & Mark, 2015; Yang, forthcoming), Russia (Nisbet et al., 2017), and the Middle Eastern monarchies (Martin et al., 2016; Wike & Simmons, 2015). *Why are many citizens in authoritarian regimes supportive of government censorship? Is it possible that participating in censorship increases citizens' support for it?*

This study takes a new step in this research field by exploring the consequences of a novel, bottom-up perspective of censorship. Instead of viewing censorship as a unilateral, top-down move by the government, I take a broader definition of censorship that includes any activities restricting public expression or access to information. This includes popular participation in state censorship, which is the focus of this research, as well as other forms of censorship not directly involving the state, such as private, peer-to-peer censorship (Beazer et al., 2022; Luo & Li, 2022; Zhao & Chen, 2023) and self-censorship (Chang & Manion, 2021; Robinson & Tannenberg, 2019; Shen & Truex, 2021; Stern & Hassid, 2012; Zhu & Fu, 2021). I hypothesize that when ordinary users actively participate in the censorship process by flagging online content they disapprove of, their support for the censorship apparatus increases. This phenomenon could be attributed to several factors, including a reduction in the perceived responsibility of the government for censorship, the creation of cognitive dissonance, and individuals motivated to justify the censorship system they participated in.

Using an original online survey in China, I first provide a novel descriptive analysis of the prevalence of public participation in the censorship process in authoritarian regimes. I find that over half of the respondents self-reported having participated in censorship. The flagged content covered a wide range of topics. While inappropriate and socially harmful content such as vulgar language is the most commonly flagged content, political discussions and tabloid gossip of entertainment stars are also widely reported. Additionally, public participation in censorship was prevalent across several demographic groups. Furthermore, I find a significant and positive correlation between participation and support for censorship.

To causally test the hypothesis, I conducted an original, pre-registered online experiment in a custom-engineered, simulated social media environment. The simulated social media page was not interactive, meaning each respondent completed the study independently. During the experiment, respondents were randomly "encouraged" to participate in censorship by reporting the social media posts they encountered. I then used an instrumental variable analysis to estimate the complier average causal effect (CACE) of the experimentally induced censorship participation. Consistent with the theory, participation in censorship significantly increased individuals' support for government censorship. I also find suggestive evidence that the institutional feature that allows public participation alone can increase support for censorship, alleviating concerns about experimenter demand effects and social desirability bias.

This study contributes to two streams of literature on authoritarian politics: government censorship and public participation. First, the study posits that censorship in authoritarian regimes should be perceived as a symbiotic relationship between the government and citizens, extending the literature that has examined authoritarian censorship solely from a top-down angle (Gallagher & Miller, 2021; Gueorguiev & Malesky, 2019; King et al., 2013; Lorentzen, 2014; Miller, 2018; Roberts, 2018; Shadmehr & Bernhardt, 2015; Yang, forthcoming). Moreover, the study provides novel empirical evidence of widespread public participation in the censorship process and explains how such participation shapes their opinions toward the government censorship apparatus. Importantly, the bottom-up perspective helps to reconcile the empirical puzzle of why repressive apparatus such as censorship continues to garner significant popular support in authoritarian regimes (Dickson, 2016; Wang & Mark, 2015; Wike & Simmons, 2015).

Second, the study extends the literature on public participation in authoritarian regimes by highlighting a novel and perhaps insidious form of public participation. While a wealth of literature has illustrated the causes and consequences of public participation in quasi-democratic institutions (Distelhorst & Hou, 2017; Gandhi, 2008; Gueorguiev, 2021; He & Warren, 2011; Manion, 2015; Stromseth et al., 2017; Truex, 2016, 2017) and

contentious social movements (Fu & Distelhorst, 2018), I show that public participation is also critical in the implementation of repressive policies like censorship (Wintrobe, 1998). Encouraging citizens' participation in the censorship process consolidates public support for authoritarian regimes' repressive apparatus. This semblance of public participation ironically suppresses individual rights to free speech and contributes to the durability of authoritarian regimes in the Internet era.

Participatory Censorship: Bottom-Up Perspective

Censorship in authoritarian regimes has traditionally been viewed as a top-down process imposed by the state and social media platforms on ordinary citizens. For example, Roberts (2018) defines censorship as "the restriction of the public expression of or public access to information *by authority* [emphasis added]." Han (2018) also regards censorship as "tools used *by the state* [emphasis added] to limit the boundaries of online expression." Recent research has expanded the purposes of government censorship beyond silencing political opposition to rewarding regime supporters (Esberg, 2020), but the consensus remains that censorship is a top-down tool of authoritarian regimes. However, this traditional view overlooks the role of ordinary citizens in censorship. This study adopts a broader definition of censorship that encompasses any activities contributing to the restriction of public expression or access to information.

This inclusive definition incorporates various types of censorship activities involving private citizens, which have not traditionally been considered censorship from a top-down perspective. First, extensive research has examined the roles of self-censorship (Chang & Manion, 2021; Robinson & Tannenberg, 2019; Shen & Truex, 2021; Stern & Hassid, 2012; Zhu & Fu, 2021) and preference falsification (Kuran, 1991) in suppressing political criticism and sustaining authoritarian regimes. Building on the logic of self-censorship, some studies have investigated how private citizens or businesses censor each other to avoid state repercussions (Beazer et al., 2022; Luo & Li, 2022; Zhao & Chen, 2023). While both self-censorship and peer censorship are primarily motivated by fear of repression, this study explores a novel perspective of censorship wherein ordinary citizens proactively participate in state censorship by reporting online content to the state and their agents (e.g., censors in social media companies).[2]

Many authoritarian regimes have a history of public participation in repressive apparatus and political campaigns (Wintrobe, 1998). During China's Cultural Revolution, for instance, ordinary citizens reported their friends, colleagues, and even families to the communist government as "counter-revolutionaries," which often led to brutal state repression (Dikötter, 2016; Jiang, 2021; Thurston, 1984; Yang 2021). This phenomenon was also

common in other dictatorships (Gregory, 2009; Wintrobe, 1998) and continues today with almost every social media platform having a function to report online content. While this feature can be used to flag inappropriate and harmful content, governments and citizens can and do abuse it for political gain (Nimmo & Agranovich, 2022). For example, Tufekci (2017) documents that Turkish Twitter users have organized mass reporting of political opponents as spam to get their accounts suspended. In China, participation in censorship is especially prevalent. Official statistics claim to receive over 172 million censorship requests in 2022 alone (Xinhua, 2023). Observers have even drawn parallels between this rising online "report culture" and the Cultural Revolution in the Mao era (BBC News, 2020; Cook, 2019; Jiang, 2021).

Although similar reporting behaviors might occur in democracies as well, authoritarian regimes, such as China, are particularly keen on encouraging citizens to participate in the censorship process and report online content (Cook, 2019; Jiang, 2021). Since the establishment of the Central Leading Group for Cybersecurity and Informatization in 2014–an entity firmly under the control of Chinese leader Xi Jinping–it has effectively assumed control over the Cyberspace Administration of China (CAC), the government agency responsible for censorship. Subsequently, encouraging public participation in censorship has emerged as a pivotal objective for the CAC. Notably, the CAC houses a dedicated division called *jubao zhongxin*, exclusively tasked with soliciting and addressing censorship appeals from ordinary users. The CAC even ran official propaganda campaigns at both central and local levels to promote public participation in censorship (CNS, 2020; Sina, 2023) and reward citizens for being "peer informants" (Cook, 2019).

Why do authoritarian governments allow and even encourage such public participation? Theories of political participation in authoritarian regimes suggest that authoritarian governments allow public participation, including in censorship, to gather valuable information such as public policy preferences and potential social unrest due to the lack of democratic institutions (Distelhorst & Hou, 2017; Gueorguiev, 2021; Manion, 2015; Stromseth et al., 2017; Truex, 2016, 2017). In the context of censorship, authoritarian governments face the challenge of identifying messages that pose a threat to the regime (Gueorguiev & Malesky, 2019; King et al., 2013). However, even in sophisticated regimes like China, complete control over the internet is difficult to achieve (Roberts, 2018). To effectively censor a large amount of information, the Chinese government reportedly employs millions of censors (King et al., 2017), uses automated keyword filtering (Han, 2018; Ng, 2015), and employs a "friction" strategy to limit access to sensitive content (Roberts, 2018).

To alleviate the information gathering problem, authoritarian governments also encourage ordinary users to participate in the censorship process (Cook,

2019; Jiang, 2021; Nimmo & Agranovich, 2022; Tufekci, 2017). The Chinese government, for instance, ran official propaganda campaigns in 2020 and 2023 to promote public participation in censorship (CNS, 2020; Cook, 2019; Sina, 2023). User reports provide valuable signals for the government and social media firms to conduct censorship, reducing the cost of monitoring the internet. Platforms like Sina Weibo establish algorithms that consider user reports when determining the publicity and censorship of posts and accounts (Cook, 2019; Jiang, 2021). By mobilizing millions of ordinary users to participate in censorship, the dynamic of censorship in China shifts from solely top-down control to a mixture of top-down control and bottom-up participation.

On the flip side, although permitting some level of free information could facilitate information gathering through bottom-up reports and stabilize the regime, allowing excessive uncontrolled information on social media could also be problematic as it can reveal widespread dissatisfaction and provide a focal point for collective actions (King et al., 2013; Lorentzen, 2014). Encouraging public participation in state censorship could create a chilling effect on ordinary citizens, leading a significant proportion to self-censor their dissatisfaction (Chang & Manion, 2021; Robinson & Tannenberg, 2019; Shen & Truex, 2021; Stern & Hassid, 2012; Zhu & Fu, 2021). Therefore, bottom-up censorship effectively alleviates the dictator's information dilemma by facilitating information gathering while inducing self-censorship, preventing public grievances from getting out of control (Wintrobe, 1998).

Participation and Public Support for Censorship

While previous studies have highlighted the negative effects of government censorship on public opinion towards the regime and its censorship apparatus (Gläßel & Paula, 2020; Pan & Siegel, 2020; Roberts, 2018, 2020), this study proposes that when ordinary citizens participate in the censorship process by reporting online content, their support for the censorship apparatus actually increases. There are several reasons to believe that public participation may have this effect.

First, public participation in censorship can diffuse the government's responsibility and create a passive image of the government. By outsourcing repression to non-state actors, the regime can plausibly deny wrongdoing and evade political accountability (Ong, 2022). In the case of censorship, the Chinese government mobilizes its pro-regime base to fabricate millions of posts to counter online critics (Chen & Xu, 2017; Han, 2015; King et al., 2017; Miller, 2018), and outsources censorship to social media platforms (Han, 2018; Miller, 2018). These tactics divert blame for censorship away from the government and onto ordinary people and social media companies. Public participation in censorship further diffuses the government's responsibility. It

creates the impression that the government is merely responding to public demand rather than initiating censorship (Luo & Li, 2022; Zhao & Chen, 2023). Therefore, even if citizens disapprove of certain censorship events, they may attribute them to other users' reports and are less likely to blame the government. Moreover, public participation creates the perception that censorship is not just the will of the government but also the will of many ordinary users, and that it is normal for online content to be removed (Yang, forthcoming).

Second, the cognitive dissonance theory, which posits that individuals experience psychological discomfort when their beliefs and behaviors are inconsistent (Festinger, 1957), might also explain the effect of public participation in increasing support. Because there is an inconsistency between reporting online content and disliking censorship, such psychological discomfort motivates individuals to reduce the inconsistencies by changing their beliefs. Thus, despite initially opposing censorship, individuals may still participate in censorship to remove content they dislike, creating a cognitive dissonance that prompts them to justify their behavior and support government censorship to reduce psychological discomfort.

Similarly, the system justification theory suggests that individuals are motivated to defend and justify the social, economic, and political systems on which they depend (Jost, 2020). In authoritarian regimes, this may lead individuals to justify initially unpleasant behaviors and experiences, such as censorship participation, to maintain a positive self-image and support for the existing regime. Furthermore, constant engagement in reporting online content may also contribute to the subconscious justification of such behavior and the censorship apparatus. Taken together, these theories support the notion that the bottom-up perspective of censorship may partly explain the popularity of the censorship apparatus in authoritarian regimes.

Hypothesis: As individuals participate more in the censorship process, they should display greater levels of support toward government censorship.

Study 1: Online Survey

Although participatory censorship is a widespread phenomenon in various autocracies (Tufekci, 2017), this study focuses on China, one of the most sophisticated censorship regimes in the world, to illustrate the features and consequences of participatory censorship. To gauge the prevalence of public participation in censorship among the Chinese population and its correlation with support for censorship, I conducted an original online survey in December 2021. The survey recruited 1124 respondents through a Chinese online survey platform, who were then directed to an anonymous survey hosted by Qualtrics, a US-based survey platform. Respondents were selected

using a quota sampling strategy to ensure a diverse range of socioeconomic backgrounds. However, like other online surveys in China (Huang, 2018; Pan et al., 2022), the sample may be younger and better educated than the general Internet population. To address this concern, the survey sample was weighted to resemble the Chinese Internet population in terms of gender, rural/urban location, region, age, and education (see Online Appendix B). In the main paper, I report results using the weighted sample and in the appendix, I report results using the original sample. The results are generally consistent.

Measurement

To measure *Participation in Censorship*, the survey asked respondents directly if they have ever reported online content or speech. Response options range from "never" to "multiple times per month," coded on a five-point scale. Although social desirability bias is a common concern in surveys conducted in authoritarian regimes, as previously discussed, participating in censorship is less taboo in China. Therefore, respondents are less likely to fear reporting their past behaviors if they have participated. Additionally, respondents are also unlikely to falsely report having participated before if they have never because there has not been any social norm to participate. Nevertheless, it is still possible that respondents might under- or over-report their participation.

For those participants who report having participated in censorship, they are asked to specify the type of content they requested to be removed from a list of options. This list includes political content, such as political news, commentary, opinions, rumors, and foreign media coverage of China. Additionally, non-political content like entertainment and advertisements are also included, as well as inappropriate content such as vulgar language and pornography.

The main outcome variable in this analysis is *Support for Censorship*. Respondents were asked whether they agreed with the statement that the government should actively control the Internet and remove content that it deems inappropriate. Additionally, I measured their *Support for Censorship of Political Content* and *Support for Censorship of Non-Political Content* by asking whether the government should or should not control online discussion of government policies and party leadership, as well as entertainment stars and popular culture. All variables were measured on a five-point scale.

Two sets of control variables were included: demographic covariates and predisposition covariates. Demographic covariates included *education*, *age*, *gender*, and *urban/rural location*. Predisposition covariates included *party membership*, *political interests*, *political ideology*, and *economic ideology*.

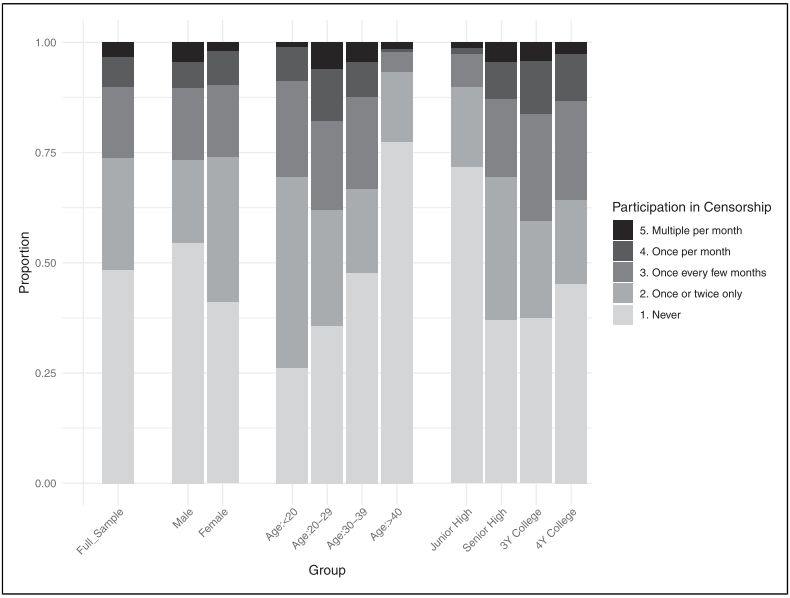

Figure 1. Distribution of self-report participation in censorship.
Note: All observations are weighted by gender, rural/urban location, region, age group, and education. The unweighted sample shows similar patterns (see Online Appendix C).

Results: Prevalence of Public Participation

How prevalent is public participation in censorship among the Chinese public? Figure 1 presents the distribution of self-reported participation in censorship using the weighted sample. As shown in the first bar on the left, more than 50% of the weighted sample report having previously participated in censorship. More than 25% have participated at least once every few months. These results demonstrate that public participation in censorship is prevalent among Chinese internet users. Such behavior is especially common among younger generations who are regularly exposed to online discussions about the use of reporting as a strategy to censor opposing opinions (Luo & Li, 2022; Zhao & Chen, 2023). Almost three-quarters of respondents under 20 years old report having such censorship experience, and around two-thirds of respondents in their 20s report similar experiences. For these young people on the Internet, flagging online content is both common and normal.

What specific content did these "participating respondents" report? The vast majority of participating respondents (around 90%) reported inappropriate content online, including pornography and vulgar language. A significant proportion of them also reported political content, with about 50% of

participating respondents (or 25% of all respondents) reporting having flagged political content. Younger and better-educated respondents were more likely to participate in political censorship, regardless of their political predisposition. Finally, around one-third of participating respondents reported censorship of entertainment and cultural content. In summary, the descriptive analysis indicates that there are significant levels of public participation in censorship across different demographic groups and categories of online content.

Participation and Censorship Support

Do individuals with higher levels of participation in censorship hold more favorable views toward the censorship apparatus? Table 1 reports the results of OLS models that investigate the relationship between participation in censorship and support for the censorship apparatus among the Chinese public. Consistent with the hypothesis, individuals who have more actively participated in the censorship process are more likely to support government censorship. Specifically, for each additional level of participation in censorship, there is a .084 increase in support for censorship on a five-point scale (8.34% of a standard deviation of the dependent variable), even after controlling for demographic and predisposition covariates. Therefore, all else equal, compared to those who never participated in censorship, individuals

Table 1. Correlation Between Participation in Censorship and Support for Censorship.

	Support for censorship		Support for Censorship of political content		Support for Censorship of non-political content	
	(1)	(2)	(3)	(4)	(5)	(6)
Participation	0.099***	0.084***	0.085***	0.105***	0.007	0.027
	(0.028)	(0.028)	(0.030)	(0.030)	(0.033)	(0.033)
Constant	2.616***	2.067***	2.549***	2.583***	2.367***	2.387***
	(0.131)	(0.182)	(0.136)	(0.191)	(0.150)	(0.211)
Demographic	✓	✓	✓	✓	✓	✓
Predisposition		✓		✓		✓
Weighted	✓	✓	✓	✓	✓	✓
N	1088	1071	1084	1066	1086	1068
Adjusted R^2	0.048	0.106	0.046	0.074	0.034	0.070

Notes: Dependent variables are indicated in column headings and are measured on a five-point scale. Standard errors are in parentheses.
*$p < .1$; **$p < .05$; ***$p < .01$

who frequently participated (multiple times per month) are, on average, 33.3% of a standard deviation higher in support of censorship. This suggests that participation in censorship is a significant predictor of support for censorship. As robustness checks, I first use alternative modeling strategies such as ordered logistic regression models. I also re-coded participation as a binary variable indicating whether the respondent has ever participated or not. The results are consistent with the main analyses (see Online Appendix D). The regression analysis results suggest that individuals who have flagged more content in the past are more likely to believe that the government should actively remove content it deems harmful or inappropriate.

The correlation between public participation in censorship and support for it also holds for political content, as evidenced by the results in columns 3 and 4 of Table 1. Respondents who have engaged more in censorship are more likely to endorse the government's active control of political news and discussions, and the magnitude of the effect is even stronger than that of the main models analyzing general support for censorship. However, there is no significant correlation between censorship participation and support for regulating non-political content.

Given the overall significant correlation between participation and support for censorship, an important theoretical question arises: is this correlation

Table 2. Correlation Between Specific Types of Participation and Support for Censorship.

	Support for censorship		Support for Censorship of political content		Support for Censorship of non-political content	
	(1)	(2)	(3)	(4)	(5)	(6)
Participation (Political)	0.059**		0.121***		0.004	
	(0.030)		(0.032)		(0.035)	
Participation (NonPolitical)		0.053		0.079**		−0.001
		(0.036)		(0.037)		(0.043)
Constant	2.151***	2.159***	2.618***	2.682***	2.435***	2.443***
	(0.178)	(0.181)	(0.187)	(0.190)	(0.207)	(0.211)
All covariates	✓	✓	✓	✓	✓	✓
Weighted	✓	✓	✓	✓	✓	✓
N	1071	1071	1066	1066	1068	1068
Adjusted R^2	0.102	0.100	0.076	0.067	0.069	0.069

Notes: *Dependent* variables are indicated in column headings and are measured on a five-point scale. Standard errors are in parentheses. The independent variables are participation in political and non-political censorship. Both demographic and predisposition covariates are included in all models.

*$p < .1$; **$p < .05$; ***$p < .01$

primarily due to the censorship behavior itself, regardless of the content being reported, or is it contingent on the content being reported? Table 2 demonstrates the correlation between specific types of participation, such as reporting political content versus non-political content, and support for censorship in general, as well as censorship of each content type.

As shown in columns 1 and 2 of Table 2, higher levels of participation in political censorship significantly correlate with higher support for censorship. Likewise, reporting entertainment and cultural content also exhibits a positive correlation with support, albeit with a lesser degree of significance (p = .138). Moreover, both types of reporting behaviors are significantly correlated with support for censorship of political content, as indicated in columns 3 and 4. These results suggest that the reporting behavior itself contributes to higher acceptance of the censorship apparatus. Even reporting non-political content might have a significant spillover effect and increase support for political censorship.

The fact that respondents who have engaged more in censorship are more likely to endorse political censorship but not non-political censorship is intriguing. After analyzing both the current observational study and the experimental study in the following section, one possible explanation is the lack of a stable and consistent public opinion toward non-political censorship. In the observational study, I used negatively worded questions asking whether respondents believe that the government should *not* restrict political/non-political discussions online. In contrast, in the experimental study, I used positively worded questions asking whether the government *should* censor. While support for political censorship is consistent across the two studies, support for non-political censorship is significantly higher when using the positive-wording question compared to the negative-wording one. This suggests a sizable confirmation bias and a lack of stable opinion toward non-political censorship. Future research should further investigate and provide more insights into non-political censorship.

In summary, the observational analyses provide evidence for the argument that public participation in censorship is widespread among the Chinese population, with approximately 50% of the respondents self-reported having engaged in censorship. Furthermore, the analysis demonstrates a positive association between participation in censorship and support for censorship, which is consistent with the main hypothesis. These findings underscore the relevance of the bottom-up perspective of censorship in authoritarian regimes. By delegating some censorship power to ordinary citizens, the Chinese government may have successfully increased public support for the censorship apparatus, contributing to the regime's stability and control over online discourse.

Mechanisms and Robustness

In the theory section, I presented three potential reasons to explain how public participation in censorship might increase support for the censorship apparatus. To examine whether participation reduces the government's responsibility for censorship, the survey asked respondents about who they believed should be accountable for censored content: netizens, the government, or platforms. Using OLS regression models with all relevant covariates and adjusted by sample weights, I find higher levels of participation in censorship are indeed significantly and negatively associated with perceived government responsibility ($\beta = -0.112$, $p = .023$). Similarly, causal mediation analysis demonstrates that perceived government responsibility has a significant and negative causal mediation effect on the relationship between participation and support for censorship.

It is more challenging to directly test the remaining two mechanisms — the cognitive dissonance theory and the system justification theory. Nonetheless, both theories imply that participation in censorship would have a more significant impact on increasing support for the censorship apparatus than increasing support for the authoritarian regime as a whole. To explore this possibility, the survey examined respondents' overall satisfaction with the Chinese regime, as well as their evaluation of the government's performance. Using OLS models, I indeed find no statistically significant relationships between participation in censorship and regime support ($\beta = -0.012$, $p = .632$). As a result, the findings suggest that participation has a stronger impact on censorship support than regime support, which provides suggestive evidence for both the cognitive dissonance theory and the system justification theory. In summary, all three mechanism arguments appear plausible for explaining the positive effect of public participation on support for censorship.

However, the observational analyses above may be susceptible to the possibility of omitted variable bias or reverse causality. To mitigate these concerns, in Online Appendix D.2, I conducted a sensitivity analysis to test the robustness of the main models to potential unobserved confounders or reverse causal relationships. The results show that the main analysis is at least robust to an omitted variable as strong as the most significant covariate, *Economic Ideology*.

It should be noted that this study does not claim that participation is exogenous to individuals' political predispositions and prior support for censorship. I acknowledge the possibility that more pro-censorship citizens might be more likely to participate. The goal of this study is to explore the downstream effects of participation in censorship on individuals' support for it. To address concerns about the correlation being driven by a reverse causal relationship and to causally test the hypothesis, I conduct an experiment that randomly manipulated individuals' participation in censorship. The following section describes the design and results of the experiment.

Study 2: Survey Experiment

Building on the first study, I conduct an original, pre-registered online survey experiment using a custom-engineered, simulated social media page. The simulated social media environment was not interactive, meaning each participant still completed the survey independently. The experiment aims to test the causal effects of participation in censorship on public support for government censorship. Since it is challenging to manipulate censorship behavior directly, I employed an instrumental variable approach, where I provided respondents in the two treatment groups with options and encouragement to report simulated social media posts. I then measured the complier average causal effect (CACE) on support for censorship among respondents who actually participated in reporting the simulated social media posts (Aronow & Carnegie, 2013; Marbach & Hangartner, 2020). This approach allows me to estimate the causal effects of censorship participation induced by the experimental treatments. Additionally, I also tested the intention-to-treat effect of the treatments by comparing the group means.

Conducting an experiment in a simulated social media environment offers several advantages over a similar field experiment on real social media. First, it avoids ethical concerns associated with field experiments, particularly given the current political climate in China, which is hostile to political research. Such an experiment might put both participants and researchers at higher risk of authoritarian repression. Furthermore, encouraging respondents to participate in censorship in the real world is normatively undesirable and might further contribute to the reporting culture on the Chinese Internet. Conducting the experiment in a simulated setting limits the potential negative impact of the research. Finally, reporting behaviors are usually not publicly observable in the real world, and in my experiment, a significant proportion of respondents object to the idea of reporting and refuse to participate in censorship even after encouragement. An experiment embedded in a simulated environment enables me to measure participation, estimate treatment effects among those who comply with the treatments, and identify those who refuse to participate regardless of treatments.

Procedure

The survey experiment was conducted in June 2022 in China. Similar to the online survey in study 1, I recruited around 4000 respondents from a Chinese online survey platform and then directed them to Qualtrics, where they completed the survey anonymously. As was the case with study 1, the sample covers a wide range of socioeconomic backgrounds but is younger and better educated than the general Internet population.

The experiment consists of three parts. First, I measured pre-treatment covariates. Second, I randomly assigned respondents to one of three groups, including a control group and two treatment groups. All participants were asked to use the simulated social media page where they read ten posts related to a hotly debated current event in China in 2022, the Xuzhou chained woman incident.[3] The ten posts were adapted from actual Sina Weibo posts with modified user names and avatars. Five of the posts are pro-government or nationalistic, while the remaining five are anti-government or pro-individual rights. The order of the posts was randomized. After reading and potentially reacting to the posts, I measured respondents' support for censorship.

On the simulated social media page, I built multiple buttons that the respondents can click under each post. In the control group, these buttons are: "Like," "Share," and "Comment." In both treatment groups 1 and 2, I built an additional "Report" button under each post that allows respondents to flag the post, mimicking the real-world participation process in censorship. Furthermore, in treatment group 2, respondents received an additional "encouragement" to use the "Report" button. Specifically, respondents in treatment group 2 were shown the following paragraph:

> We are especially interested in what posts you want to report. Please choose at least two posts that you think should be removed by the Internet regulator, and press the Report button to let us know.[4]

Table 3 summarizes the experimental design and treatment assignments. By comparing the difference in means between the control group and treatment group 1, we can test the impact of the institutional feature that permits public participation in censorship. Although this is not a direct test of the main hypothesis regarding the effects of participation behavior, it offers additional insights into the bottom-up perspective of authoritarian censorship. If we observe significant effects from merely providing the institutional feature for public participation, it may suggest that the effect sizes would be even more substantial if respondents actually participated in the censorship process. Furthermore, having both treatments mitigates the

Table 3. Summary of Experimental Design and Treatments.

Groups	Control	Treatment 1	Treatment 2
Buttons under simulated posts	Like, share, comment	Like, share, comment, **report**	Like, share, comment, **report**
Encouragement Message	No	No	Yes

concerns of experimenter demand effects and social desirability bias. Since respondents in treatment group 1 were not explicitly instructed to participate in censorship, they were less likely to infer the objectives of the experiment or the social norms around censorship participation, reducing the likelihood of changing their behavior and providing socially desirable answers.

To measure respondents' participation, I use a binary variable indicating whether they clicked any of the "Report" buttons. As in the previous study, the main dependent variable is *Support for Censorship*, and I also measure *Support for Censorship of Political Content* and *Support for Censorship of Non-political Content*. To ensure balance across the three experimental groups, I include ten pre-treatment covariates in the analysis. Among these covariates, four are demographic variables, including *education*, *age*, *gender*, and *region*, which are commonly used in experiments across various contexts. The remaining six covariates are predisposition covariates, including *party membership*, *nationalism*, *political interests*, *ideology*, *social media usage*, and *foreign connection*. All covariates are balanced across the three experimental groups.

Results: Difference in Means

I first report the overall results of the experiment by comparing the group means of the outcome variables across the three groups and then elaborate on the instrumental variable analysis that more precisely identifies the causal effect of participation on support for censorship. In the control group, none of the respondents reported any simulated posts as they were not allowed to do so. In treatment group 1, 43% of the respondents clicked the "Report" buttons on the simulated social media page, while 64% of the respondents in treatment group 2 did so. Thus, both treatments successfully induced participation among respondents who would not have otherwise participated.

Figure 2 presents the difference in means using the control group as the reference group. The left bar in each panel compares the control group and treatment group 1, and the results suggest that respondents who were given the opportunity to report simulated posts expressed higher support for censorship in general ($\beta = 0.067$, $p = .065$) and censorship of political content ($\beta = 0.066$, $p = .076$). Both results are not statistically significant at the conventional level but significant at the 0.1 level. However, I did not find a significant difference in support for censorship of non-political content, possibly due to the political nature of the selected topic. These findings suggest that simply providing the institutional feature to flag online content, without explicit messages that might induce experimenter demand effects or social desirability bias, can potentially increase individuals' support for censorship.

Moving on to treatment group 2, the right bars in each panel demonstrate that respondents who were given the option to report and encouraged to flag

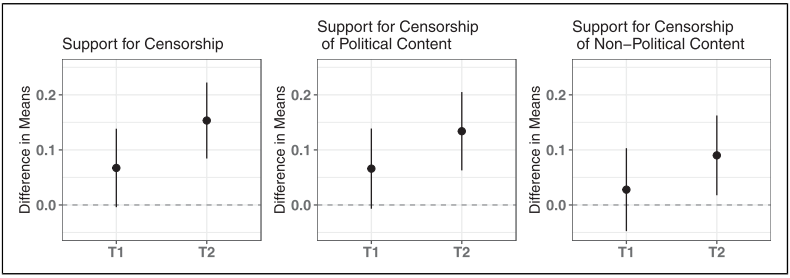

Figure 2. Difference in means of the outcome variables.
Note: All three outcome items are measured on a five-point scale. The control groups are the reference groups and bars indicate 95% confidence intervals.

online content showed significantly higher levels of support for government censorship in general ($\beta = 0.153$, $p < .001$), as well as support for censorship of political content ($\beta = 0.134$, $p < .001$) and non-political content ($\beta = 0.090$, $p = .015$) in particular. The effect sizes are larger than treatment group 1, indicating that the additional participation in censorship induced by the encouragement message further increases support for the censorship apparatus.

All in all, the overall results of the experiment provide strong and consistent evidence for the central argument that increased participation in censorship leads to higher support for the censorship apparatus. The institutional feature that allows public participation and the encouragement treatment that directly increases reporting behaviors both generate significant support for censorship, giving us greater confidence that the increase in support is not merely an experimenter demand effect or due to social desirability bias. However, it is important to note that neither treatment directly measures reporting behavior. To estimate more precisely the effect of reporting behaviors on support for censorship, the next section introduces the instrumental variable analysis and reports its findings.

Instrumental Variable Analysis

To directly identify the effect of participation in censorship, I use both treatments as instruments to estimate the complier average causal effect (CACE) of participation in censorship (clicking the report button) on support for censorship. Formally:

$$\text{Clicking the Report Button}_i = \alpha + \gamma \cdot \text{Treatment Group} + \lambda Z_i + \epsilon_i$$

$$Y_i = \xi + \beta \cdot \widehat{\text{Clicking the Report Button}}_i + \delta Z_i + \mu_i$$

where Y_i is the outcome measure; Z_i is a vector of pre-treatment covariates; and β is the CACE. In the instrumental variable analysis, I included all three groups and treated the two treatment conditions as factors that independently influence participation in censorship. In Online Appendix F.2, I present several alternative instrumental variable models, including (1) using an ordinal variable for the two levels of treatment and (2) analyzing the data with only two of the three groups. The results from alternative modeling strategies do not alter the substantive interpretation of the main results.

Table 4 reports the results from instrumental variable analyses and the CACE of participating in censorship in the simulated social media environment. Consider, first, column 2. After controlling for pre-treatment covariates, participation induced by the encouragement treatment significantly increases respondents' general support for government censorship ($\beta = 0.219$, $p < .001$). This again provides direct support for the main hypothesis and more importantly, it addresses the concerns in Study 1 that the causal arrow might be reversed. The magnitude of the treatment effect on support for censorship is considerable, equivalent to 25% of a standard deviation. This is a substantial increase given that the baseline support for censorship is already high in the control group. The instrumental variable analysis also indicates that censorship participation induced by both treatments increases specific support for censorship of political content ($\beta = 0.199$, $p < .001$) and non-political content ($\beta = 0.123$, $p = .029$), further supporting the theoretical expectations.

I further disaggregate the CACE for Treatments 1 and 2 in Online Appendix F. The results show that, while both treatments significantly

Table 4. Complier Average Causal Effects (CACE) of Participating in Censorship on Support for Censorship.

	Support for Censorship		Support for censorship of political content		Support for censorship of non-political content	
	(1)	(2)	(3)	(4)	(5)	(6)
ReportClick	0.228***	0.219***	0.202***	0.199***	0.130**	0.123**
	(0.054)	(0.052)	(0.055)	(0.054)	(0.058)	(0.056)
Constant	3.484***	1.650***	3.463***	1.909***	3.673***	1.931***
	(0.024)	(0.121)	(0.025)	(0.125)	(0.025)	(0.130)
Covariates		✓		✓		✓
N	3990	3764	3997	3770	3989	3763

Notes: Dependent variables are indicated in column headings and are measured on a five-point Likert scale. Standard errors in parentheses.
Report click is a binary variable indicating whether the respondents have clicked any of the "Report" buttons on the simulated social media page.
*$p < .1$; **$p < .05$; ***$p < .01$.

increased support through inducing reporting behaviors, the effects of the explicit instruction (Treatment 2) are more than twice as large as those of simply providing the institutional feature (Treatment 1). This suggests that some respondents may have self-censored their disapproval when explicitly primed with bottom-up censorship, indicating that participation might stabilize authoritarian regimes through a combination of increasing sincere support and inducing self-censorship. Additionally, I analyze respondents' reporting behavior based on whether they clicked "report" under a pro-government post or an anti-government post (see Online Appendix F.2). Unsurprisingly, the majority of reports were directed toward anti-regime posts. However, reporting both pro- and anti-regime posts increased support for the censorship apparatus. These results highlight the efficacy of the bottom-up censorship strategy in cultivating support across different socio-political groups.

To check the robustness of the treatment effect, I employ an alternative measure of participation in censorship and re-run the instrumental variable analyses (see Online Appendix F.2). Instead of a binary variable indicating whether the respondents clicked any of the "Report" buttons, I use the count of times respondents clicked the "Report" button. The results from this analysis are consistent with the main analyses, indicating that the additional clicking of the "Report" button induced by the encouragement treatment leads to a significant increase in support for censorship. Additionally, I have subset the data to estimate the effect of participation induced by the option to report and the effect of participation induced by the encouragement message separately. Both analyses reveal a significant CACE of reporting simulated posts on the support of censorship.

While the instrumental variable analyses indicate significant and positive treatment effects of censorship participation on support for government censorship, it is important to compare the background characteristics of compliers, those who participated when given the opportunity, and non-compliers, those who did not even when the opportunity was present, to comprehend which demographic groups drive the results (Marbach & Hangartner, 2020). In treatment group 1, individuals who clicked the "Report" buttons tended to be younger and more familiar with social media. In treatment group 2, the additional participants due to the encouragement message were also young, well-educated, nationalistic, and had more foreign connections. In contrast, the non-compliers, those who never click the "Report" buttons, tended to be older, less nationalistic, and had limited social media exposure (see Online Appendix F.3 for more details). Therefore, the results of the instrumental analyses are primarily driven by younger generations who tend to be more educated and familiar with social media.

There are several potential explanations for why participatory censorship is most pronounced among young Chinese individuals. One prominent reason

might be technological literacy. As shown in Study 1, younger respondents are more familiar with social media and thus more likely to engage in censorship activities, while older counterparts are less knowledgeable about these social media functions. However, a more deep-rooted explanation is the noticeable shift away from liberalism among the youngest generation in China. As Pan and Xu (2018) demonstrates, liberal values peaked among Millennials and began to decline with Gen Z, potentially explaining the wider acceptance of illiberal activities such as bottom-up censorship. Thus, it is less certain whether there are comparable effects of participation on support for censorship among older individuals. However, the composition of compliers and non-compliers may differ depending on the chosen topic and posts, and the profiling outcomes should be interpreted with caution. Future research should test the hypothesis in various contexts to improve the generalizability of the findings.

Implications

To sum up the findings in the experiment, both the institutional feature that allows public participation in censorship and the message that directly encourages participation increase support for censorship to various degrees. The instrumental variable analysis further demonstrates that the treatment effects are large and significant among respondents who actually participated in reporting the simulated social media posts.

The implications of the experimental results are three-fold. First, they echo the observations made in the previous survey that public participation in the censorship process is a widespread phenomenon in authoritarian China. As demonstrated by the treatment groups, approximately 40% of the respondents engaged in censorship when given the chance, and a simple encouragement message resulted in nearly two-thirds of the respondents participating in censorship.

Second, the results reaffirm the positive relationship between widespread participation in the censorship process and public support for censorship, highlighting the importance of a comprehensive understanding of authoritarian censorship that combines top-down control and bottom-up participation. The institutional feature that permits public participation and the encouragement message that directly increases reporting behaviors can both generate significant support for the censorship apparatus. Furthermore, the results are unlikely to be merely an experimenter demand effect since respondents in treatment group 1 expressed higher support even without explicit instructions to flag online content.

Third, the profiling of compliers indicates that the effect is mainly driven by young, well-educated individuals with foreign connections. This group is traditionally perceived as less vulnerable to authoritarian controls and more

receptive to Western values (Huang, 2015). The findings suggest that authoritarian regimes can leverage public participation in repressive apparatus such as censorship as a means to generate support for its repressive apparatus among this young and educated demographic, thereby maintaining popular support across generations. Overall, the experiment sheds light on the symbiotic relationship between the authoritarian government and its citizens in the censorship process, particularly the bottom-up perspective that existing studies tend to overlook, and provides causal evidence for the positive effects of public participation on support for repressive apparatus such as censorship.

One important limitation of the current study is the inability to track individuals' downstream political behaviors on the Internet. Prior research on state censorship and Internet behaviors has demonstrated that individuals encountering top-down state censorship may engage in self-censorship under certain circumstances while expressing backlash in other situations (Roberts, 2018; Zhu & Fu, 2021). Future research should further investigate whether similar theories apply to bottom-up censorship behaviors, providing valuable insights into the self-regulation mechanisms triggered by the reporting feature.

Conclusion

Repressive institutions such as censorship have long been viewed as instruments of top-down suppression employed by authoritarian regimes to quash political opposition. This study offers a new perspective by examining the prevalence and significance of bottom-up public participation in censorship in authoritarian regimes. Through two original online surveys conducted in China, one observational and one experimental, I demonstrate that public participation in censorship is a pervasive phenomenon that significantly shapes public opinion towards the censorship apparatus. Therefore, censorship in authoritarian regimes should be viewed as a symbiotic relationship between the government and its citizens, echoing the longstanding history of public involvement in repressive apparatus and political campaigns, such as the Cultural Revolution.

The findings in this study highlight the discrepancy between the common understanding of repressive authoritarian apparatus, such as censorship, in the Western world and how ordinary citizens in authoritarian regimes perceive and interact with these repressive apparatuses. For many citizens in China, censorship and other repressive institutions have been normalized as part of the rules of political life (Yang, forthcoming). Rather than fighting against the rules, individuals in these regimes take advantage of the censorship apparatus to suppress opposing views (Luo & Li, 2022; Tufekci, 2017; Zhao & Chen, 2023). Consequently, citizens no longer view the regime as the oppressor, but rather as a powerful arbitrator of censorship demands that they must win over in their internal struggles against fellow citizens.

Beyond autocracies, censorship has become an important social issue in many democracies including the United States, and public participation in censorship has also become more prevalent. Indeed, despite major differences in political systems, Appel et al. (2023) found a similar pattern in Americans' willingness to report online content and their attitudes toward censorship, suggesting the generalizability of this study. American social media platforms such as Twitter have also introduced community-based bottom-up content moderation projects that involve public participation. Although democracies might care about different policy implications, such as electoral integrity, compared with their authoritarian counterparts, it is still important to examine the consequences of these content moderation projects, because the balancing act of fighting against misinformation and preserving freedom of speech is difficult yet critical for sustaining a healthy democracy.

Acknowledgements

I thank Margit Tavits, Deniz Aksoy, Taylor Carlson, Carly Wayne, Haifeng Huang, Ted Enamorado, Jeremy Siow, Taylor Damann, Dahjin Kim, Rex Weiye Deng, other colleagues at Washing University in St. Louis, and seminar participants at MPSA, UC San Diego, and Nuffield College for their helpful comments and suggestions. I am grateful for the financial support from the Centennial Center of the American Political Science Association, the Weidenbaum Center at Washington University in St. Louis, and the Chiang Ching-Kuo Foundation in Taiwan.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study is supported by the Centennial Center for Political Science and Public Affairs of the American Political Science Association and the Weidenbaum Center on the Economy, Government, and Public Policy at Washington University in St. Louis and Chiang Ching-Kuo Foundation for International Scholarly Exchange.

ORCID iD

Tony Zirui Yang https://orcid.org/0000-0001-5948-5354

Data Availability Statement

Supplementary material is available online. Replication materials and codes can be found in the CPS Dataverse (Yang, 2024).

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Replication files are available in the CPS Dataverse (Yang 2024). Both studies were approved by the Institutional Review Board at Washington University in St Louis and pre-registered on the Open Science Framework.
2. A broad definition of censorship might also include normative positive activities such as fact-checking rumors and restricting hate speech. However, recent research in both China (Yang, forthcoming) and the United States (Appel et al., 2023) demonstrates that these activities can significantly influence public perceptions of censorship or content moderation. Additionally, as detailed in the empirical section, over 90% of respondents who reported participating in censorship indicated that they reported "inappropriate content." This participation not only increases their support for censorship as a whole but also for political censorship.
3. In January 2022, a video of a trafficked woman held in chains in a hut in Fengxian County, Xuzhou City for years went viral. Government officials were heavily criticized on social media for causing such tragedy and, more importantly, trying to cover it up. However, the incident coincided with the 2022 Beijing Winter Olympics. Such timing prompted many patriotic regime supporters to argue that this was a conspiracy to defame China.
4. Although respondents were encouraged to report at least two posts, they were not forced to do so. Non-compliance was allowed.

References

Appel, R. E., Pan, J., & Roberts, M. E. (2023). Partisan conflict over content moderation is more than disagreement about facts. *Science Advances*, *9*(44), Article eadg6799. https://doi.org/10.1126/sciadv.adg6799

Aronow, P. M., & Carnegie, A. (2013). Beyond LATE: Estimation of the average treatment effect with an instrumental variable. *Political Analysis*, *21*(4), 492–506. https://doi.org/10.1093/pan/mpt013

BBC News. (2020). *China where Fang Fang is from: The report culture is spreading along with the rise of nationalism*. BBC News Chinese Version. https://www.bbc.com/zhongwen/simp/chinese-news-52560990

Beazer, Q. H., Crabtree, C. D., Fariss, C. J., & Kern, H. L. (2022). When do private actors engage in censorship? Evidence from a correspondence experiment with Russian private media firms. *British Journal of Political Science*, *52*(4), 1790–1809. https://doi.org/10.1017/s0007123421000351

Chang, C., & Manion, M. (2021). Political self-censorship in authoritarian states: The spatial-temporal dimension of trouble. *Comparative Political Studies*, *54*(8), 1362–1392. https://doi.org/10.1177/0010414021989762

Chang, K.-C., Hobbs, W. R., Roberts, M. E., & Steinert-Threlkeld, Z. C. (2022). COVID-19 increased censorship circumvention and access to sensitive topics in China. *Proceedings of the National Academy of Sciences*, *119*(4), Article e2102818119. https://doi.org/10.1073/pnas.2102818119

Chen, J., & Xu, Y. (2017). Why do authoritarian regimes allow citizens to voice opinions publicly? *The Journal of Politics*, *79*(3), 792–803. https://doi.org/10.1086/690303

Chen, Y., & Yang, D. Y. (2019). The impact of media censorship: 1984 or brave new world? *The American Economic Review*, *109*(6), 2294–2332. https://doi.org/10.1257/aer.20171765

CNS. (2020). *2020 Shanghai shi wangluo jubao xuanchuan Yue zhengshi qidong [2020 Shanghai internet reporting promotion month officially launched]*. China News Agency. https://www.chinanews.com/sh/2020/09-02/9280670.shtml

Cook, S. (2019). *Analysis: How the Chinese communist party is incentivizing re-pression* (Vol. *133*). China Media Bulletin.

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

Dikötter, F. (2016). *The cultural revolution: A people's history, 1962–1976*. Bloomsbury Publishing USA.

Distelhorst, G., & Hou, Y. (2017). Constituency service under nondemocratic rule: Evidence from China. *The Journal of Politics*, *79*(3), 1024–1040. https://doi.org/10.1086/690948

Esberg, J. (2020). Censorship as reward: Evidence from pop culture censorship in Chile. *American Political Science Review*, *114*(3), 821–836. https://doi.org/10.1017/s000305542000026x

Festinger, L. (1957). . *A theory of cognitive dissonance* (Vol. *2*). Stanford University Press.

Fu, D., & Distelhorst, G. (2018). Grassroots participation and repression under Hu Jintao and Xi Jinping. *The China Journal*, *79*(1), 100–122. https://doi.org/10.1086/694299

Gallagher, M., & Miller, B. (2021). Who not what: The logic of China's information control strategy. *The China Quarterly*, *248*(1), 1011–1036. https://doi.org/10.1017/s0305741021000345

Gandhi, J. (2008). *Political institutions under dictatorship*. Cambridge University Press.

Gläßel, C., & Paula, K. (2020). Sometimes less is more: Censorship, news falsification, and disapproval in 1989 East Germany. *American Journal of Political Science*, *64*(4), 682–698. https://doi.org/10.1111/ajps.12501

Gregory, P. R. (2009). *Terror by quota*. Yale University Press.

Gueorguiev, D. (2021). *Retrofitting Leninism: Participation without democracy in China*. Oxford University Press.

Gueorguiev, D. D., & Malesky, E. J. (2019). Consultation and selective censorship in China. *The Journal of Politics*, *81*(4), 1539–1545. https://doi.org/10.1086/704785

Han, R. (2015). "Defending the authoritarian regime online: China's" voluntary fifty-cent army. *The China Quarterly*, *224*, 1006–1025. https://doi.org/10.1017/s0305741015001216

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

He, B., & Warren, M. E. (2011). Authoritarian deliberation: The deliberative turn in Chinese political development. *Perspectives on Politics*, *9*(2), 269–289. https://doi.org/10.1017/s1537592711000892

Hobbs, W. R., & Roberts, M. E. (2018). How sudden censorship can increase access to information. *American Political Science Review*, *112*(3), 621–636. https://doi.org/10.1017/s0003055418000084

Huang, H. (2015). International knowledge and domestic evaluations in a changing society: The case of China. *American Political Science Review*, *109*(3), 613–634. https://doi.org/10.1017/s000305541500026x

Huang, H. (2018). The pathology of hard propaganda. *The Journal of Politics*, *80*(3), 1034–1038. https://doi.org/10.1086/696863

Jiang, J. (2021). The eyes and ears of the authoritarian regime: Mass reporting in China. *Journal of Contemporary Asia*, *51*(5), 828–847. https://doi.org/10.1080/00472336.2020.1813790

Jost, J. T. (2020). *A theory of system justification*. Harvard University Press.

King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review*, *107*(2), 326–343. https://doi.org/10.1017/s0003055413000014

King, G., Pan, J., & Roberts, M. E. (2017). How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. *American Political Science Review*, *111*(3), 484–501. https://doi.org/10.1017/s0003055417000144

Kuran, T. (1991). Now out of never: The element of surprise in the East European revolution of 1989. *World Politics*, *44*(1), 7–48. https://doi.org/10.2307/2010422

Lorentzen, P. (2014). China's strategic censorship. *American Journal of Political Science*, *58*(2), 402–414. https://doi.org/10.1111/ajps.12065

Luo, Z., & Li, M. (2022). Participatory censorship: How online fandom community facilitates authoritarian rule. *New Media & Society*, *26*(2), 4236–4254. https://doi.org/10.1177/14614448221113923

Manion, M. (2015). *Information for autocrats: Representation in Chinese local congresses*. Cambridge University Press.

Marbach, M., & Hangartner, D. (2020). Profiling compliers and noncompliers for instrumental-variable analysis. *Political Analysis*, *28*(3), 435–444. https://doi.org/10.1017/pan.2019.48

Martin, J. D., Martins, R. J., & Wood, R. (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), 3400–3422.

Miller, B. (2018). Delegated dictatorship: Examining the state and market forces behind information control in China PhD thesis.

Ng, J. Q. (2015). *Politics, rumors, and ambiguity: Tracking censorship on WeChat's public accounts platform*. Munk School of Global Affairs.

Nimmo, B., & Agranovich, D. (2022). Meta's adversarial threat report, First Quarter 2022.

Nisbet, E. C., Kamenchuk, O., & Dal, A. (2017). A psychological firewall? Risk perceptions and public support for online censorship in Russia. *Social Science Quarterly*, *98*(3), 958–975. https://doi.org/10.1111/ssqu.12435

Ong, L. H. (2022). *Outsourcing repression: Everyday state power in contemporary China*. Oxford University Press.

Pan, J., Shao, Z., & Xu, Y. (2022). How government-controlled media shifts policy attitudes through framing. *Political Science Research and Methods*, *10*(2), 317–332. https://doi.org/10.1017/psrm.2021.35

Pan, J., & Siegel, A. A. (2020). How Saudi crackdowns fail to silence online dissent. *American Political Science Review*, *114*(1), 109–125. https://doi.org/10.1017/s0003055419000650

Pan, J., & Xu, Y. (2018). China's ideological spectrum. *The Journal of Politics*, *80*(1), 254–273. https://doi.org/10.1086/694255

Pop-Eleches, G., & Way, L. A. (2021). Censorship and the impact of repression on dissent. *American Journal of Political Science*, *67*(2), 456–471. https://doi.org/10.1111/ajps.12633

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

Roberts, M. E. (2020). Resilience to online censorship. *Annual Review of Political Science*, *23*(1), 401–419. https://doi.org/10.1146/annurev-polisci-050718-032837

Robinson, D., & Tannenberg, M. (2019). Self-censorship of regime support in authoritarian states: Evidence from list experiments in China. *Research & Politics*, *6*(3), Article 2053168019856449. https://doi.org/10.1177/2053168019856449

Shadmehr, M., & Bernhardt, D. (2015). State censorship. *American Economic Journal: Microeconomics*, *7*(2), 280–307. https://doi.org/10.1257/mic.20130221

Shen, X., & Truex, R. (2021). In search of self-censorship. *British Journal of Political Science*, *51*(4), 1672–1684. https://doi.org/10.1017/s0007123419000735

Sina. (2023). *Zhong yang Wang Xin Ban Zhao Kai Quan Guo Wang Luo ju bao gong zuo Hui Yi [The Central Cyberspace Administration of China Held a National Online Reporting Work Meeting]*. Sina News. https://finance.sina.com.cn/tech/roll/2023-04-18/doc-imyqumiq8779068.shtml

Stern, R. E., & Hassid, J. (2012). Amplifying silence: Uncertainty and control parables in contemporary China. *Comparative Political Studies*, *45*(10), 1230–1254. https://doi.org/10.1177/0010414011434295

Stromseth, J. R., Malesky, E. J., Gueorguiev, D. D., Lai, H., & Brinton, C. (2017). *China's governance puzzle: Enabling transparency and participation in a single-party state*. Cambridge University Press.

Thurston, A. F. (1984). Victims of China's cultural revolution: The invisible wounds: Part I. *Pacific Affairs*, *57*(4), 599–620. https://doi.org/10.2307/2758711

Truex, R. (2016). *Making autocracy work: Representation and responsiveness in modern China*. Cambridge University Press.

Truex, R. (2017). Consultative authoritarianism and its limits. *Comparative Political Studies*, *50*(3), 329–361. https://doi.org/10.1177/0010414014534196

Tufekci, Z. (2017). *Twitter and tear gas*. Yale University Press.

Wang, D., & Mark, G. (2015). Internet censorship in China: Examining user awareness and attitudes. *ACM Transactions on Computer-Human Interaction*, *22*(6), 1–22. https://doi.org/10.1145/2818997

Wike, R., & Simmons, K. (2015). *Global support for principle of free expression, but opposition to some forms of speech* (Vol. *18*) Pew Research Center.

Wintrobe, R. (1998). *The political economy of dictatorship*. Cambridge University Press.

Xinhua. (2023). *2022 Nian quanguo shouli wangluo weifa he buliang xinxi jubao 1.72 Yi Jian [In 2022, 172 million reports of illegal and inappropriate online content nationwide]*. Xinhua News Agency. https://www.news.cn/politics/2023-02/10/c_1129353484.htm

Yang, J. (2021). *The world turned upside down: A history of the Chinese cultural revolution*. Farrar, Straus and Giroux.

Yang, T. Z. (2024). *Replication data for: Participatory censorship in authoritarian regimes*. Harvard Dataverse. https://doi.org/10.7910/DVN/P8YWA4

Yang, T. Z. forthcoming. Normalization of censorship: Evidence from China. *The Journal of Politics*.

Zhao, A., & Chen, Z. (2023). Let's report our rivals: How Chinese fandoms game content moderation to restrain opposing voices. *Journal of Quantitative Description: Digital Media*, *3*. https://doi.org/10.51685/jqd.2023.006

Zhu, Y., & Fu, K.-W (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), 3634–3655. https://doi.org/10.1177/1461444820959016

Author Biography

Tony Zirui Yang is a Postdoctoral Prize Research Fellow at Nuffield College, University of Oxford, an Associate at University of Oxford China Centre, and an incoming Assistant Professor of Political Science at Emory University. His work focuses on authoritarian politics, political communication, and public opinion in authoritarian regimes.