Normalization of Censorship: Evidence from China
Online Appendices

February 23, 2023

Contents

Appendices

Appendix A   Experiments: Survey Procedure and Descriptive Statistics

A.1   Survey Procedure & Pre-Registration

The first survey experiment was conducted in December 2020. The second survey experiment was conducted in December 2022. Both surveys were administered in mainland China by a Shanghai-based Chinese online survey company. The participants were recruited by the survey company and then directed to a US-based website, Qualtrics, where they completed the survey anonymously. Once they completed the survey on Qualtrics, they were redirected back to the survey vendor's platform.

All mainland Chinese citizens above 18 years old are eligible for this study. To make sure that the sample covers a broad range of socioeconomic backgrounds, I put quotas on gender, region, education, and age. In the end, the quotas successfully yielded samples that reflect the population in terms of gender and region. The age distributions are also pretty close to the demographic considering the fact that younger people under 18 are not eligible for the study. The education quotas alleviate the problem of homogeneous survey participants but fall short of yielding a sample representative of the Internet population.

To further ensure sample quality, I used attention checks to screen the respondents at the beginning of the surveys. About 60% of the respondents passed the attention checks yielding 612 valid responses in study 1 and 3,314 valid responses in study 2.

A.2   Compliance with Ethical Principles of Human Subject Research

Both surveys followed all established principles of human subject research and were approved by the Institutional Review Board (IRB) at the researcher's home institution. Although the IRB exempted both studies from a formal consent form, I still included a consent page and information sheet at the beginning of both surveys. All participants were informed about the purpose, length, and format of the study. All participants need to click "I consent" on the information sheet page before they could proceed. They were allowed to opt out of the study at any point in the survey. Incomplete survey responses were not recorded.

Because the treatment prompt explicitly asked the respondents to **imagine** that they were reading WeChat articles, no deception was used. All articles in both experiments

were actual WeChat articles that were censored by WeChat. At the end of both surveys, participants were explicitly told that this was an experimental study and information in the survey might not be representative of reality.

All respondents were paid by the survey firm at its usual rate for their participation. The survey firm was paid by the researcher of this study. All participants were adults and none of them would be put in a disadvantageous position had they chosen not to participate.

Because both surveys were conducted in China, an authoritarian regime, I paid extra caution to protect respondents' information and responses, so that they will not be negatively affected by the authority due to their participation in this study. I did not ask for personal information that could directly identify participants' identities, such as names, phone numbers, and email addresses. I stored all the responses at Qualtrics via an American institutional account. The study passed the information security review at the researcher's home institution.

A.3 Survey Sample

Table A1: Sociodemographics of the Study Participants and Chinese Internet Users

	Sociodemographics	Study 1	Study 2	Chinese Internet Users
Region	East	50.8%	54.8%	46.2%
	Northeast	7.8%	6.0%	8.4%
	Central	19.6%	21.7%	22.1%
	West	21.6%	17.6%	23.3%
Gender	Female	50.0%	50.1%	48.1%
	Male	50.0%	49.9%	51.9%
Education	≤ Junior high	3.6%	3.7%	56.1%
	Senior high	12.6%	16.5%	23.8%
	3-year college	25.2%	36.4%	10.5%
	≥ 4-year college	58.5%	43.3%	9.7%
Age	≤ 19	6.4%	2.2%	23.2%
	20-29	31.4%	27.4%	21.5%
	30-39	45.1%	46.2%	20.8%
	40-49	14.9%	15.9%	17.6%
	≥ 50	2.1%	8.3%	16.9%
Income	≤ 3000	7.8%	6.4%	51.0%
	3000-5000	13.9%	13.8%	21.5%
	5000-8000	38.2%	32.6%	14.3%
	≥ 8000	38.9%	47.2%	13.3%
Occupation	Student	8.3%		26.9%
	Self-employed	13.1%		22.4%
	Corporate employee	34.5%		8.0%
	Corporate management	16.3%		2.9%
	Government employee	2.8%		2.8%
	Professional	12.6%		6.0%
	Manufacturing	4.2%		2.6%
	Service worker	3.6%		4.4%
	Migrant worker	2.0%		4.2%
	Farmer	0.7%		6.3%
	Unemployed & Retired	2.0%		13.5%
Location	Urban	71.9%		71.8%
	Rural	28.1%		28.2%

Note: Data about Chinese Internet users are from *The 45th Statistical Report of Internet Development in China*, issued by China Internet Network Information Center in April 2020.

A.4 Balance Table

Table A2: Balance Table

	Study 1			Study 2			Combined		
	Control	Treated	p	Control	Treated	p	Control	Treated	p
Female	0.469	0.531	.12	0.484	0.507	.28	0.481	0.513	.10
Age Group	3.863	3.931	.59	4.447	4.342	.16	4.318	4.252	.31
Education	3.407	3.447	.58	3.198	3.196	.96	3.244	3.251	.83
Income	3.221	3.242	.82	3.383	3.338	.34	3.348	3.317	.47
Party Member	0.247	0.274	.46	0.136	0.124	.38	0.161	0.156	.74
Ideology	2.668	2.541	.23	2.363	2.290	.07	2.431	2.344	.03
Pol Interest	4.121	4.085	.74	3.781	3.751	.53	3.856	3.824	.48
Social Media	3.313	3.398	.30	3.523	3.497	.53	3.476	3.476	.99

Table A3: Using Covariates to Predict Treatment

	Treatment		
	Study 1	Study 2	Combined
Female	0.050	0.023	0.030
	(0.042)	(0.022)	(0.020)
Education	0.018	0.0003	0.005
	(0.028)	(0.017)	(0.014)
Age Group	0.013	−0.009	−0.005
	(0.015)	(0.007)	(0.006)
Income	0.002	−0.008	−0.005
	(0.024)	(0.012)	(0.011)
Ideology	−0.012	−0.023	−0.020**
	(0.016)	(0.012)	(0.009)
Party Member	0.032	−0.015	−0.002
	(0.050)	(0.033)	(0.027)
Political Interest	−0.018	−0.004	−0.005
	(0.018)	(0.010)	(0.009)
Social Media Usage	0.020	−0.005	0.001
	(0.021)	(0.012)	(0.010)
N	593	2,119	2,712

**p < .05; ***p < .01

As shown in Table A2 and Table A3, the randomization in general is successful, producing mostly balanced groups. However, there is a slight imbalance in ideology, likely due to chance.

Appendix B Experiments: Additional Analyses

B.1 OLS Regressions with Covariates

In this section, I report regression results with all pre-treatment covariates for studies 1 and 2. For the combined sample, because one of the pre-treatment covariates is imbalanced (see Table A2 and Table A3), I report regression results for both controlling that imbalanced variable only and all pre-treatment covariates. The results are mostly consistent with the main results reported in the main paper.

Table B1: Treatment Effects on Support for the Censorship Apparatus

| | Support for Censorship Apparatus | | | |
	Study 1	*Study 2*	*Combined*	*Combined*
Treatment	0.213***	0.146***	0.155***	0.162***
	(0.075)	(0.040)	(0.035)	(0.035)
Female	0.121	0.121***		0.128***
	(0.076)	(0.041)		(0.036)
Education	−0.019	−0.034		−0.027
	(0.050)	(0.031)		(0.026)
Age Group	0.016	0.020		0.013
	(0.028)	(0.013)		(0.011)
Income	0.103**	0.106***		0.098***
	(0.044)	(0.022)		(0.019)
Ideology	−0.325***	−0.316***	−0.326***	−0.320***
	(0.029)	(0.022)	(0.017)	(0.017)
Party Member	0.109	−0.151**		−0.054
	(0.090)	(0.060)		(0.049)
Political Interest	−0.038	0.028		0.022
	(0.034)	(0.019)		(0.016)
Social Media Usage	−0.013	−0.006		−0.012
	(0.038)	(0.021)		(0.018)
Constant	4.003***	3.675***	4.172***	3.764***
	(0.247)	(0.159)	(0.048)	(0.134)
N	584	2,088	2,733	2,672
Adjusted R^2	0.203	0.129	0.127	0.143

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

Table B2: Treatment Effects on Regime Support: Overall Satisfaction

| | Regime Support: Overall Satisfaction | | | |
	Study 1	*Study 2*	*Combined*	*Combined*
Treatment	0.194***	0.086**	0.114***	0.112***
	(0.070)	(0.038)	(0.033)	(0.033)
Female	0.111	0.123***		0.131***
	(0.071)	(0.039)		(0.034)
Education	0.025	−0.029		−0.004
	(0.046)	(0.029)		(0.025)
Age Group	0.016	−0.038***		−0.035***
	(0.026)	(0.012)		(0.011)
Income	0.017	0.086***		0.069***
	(0.040)	(0.021)		(0.018)
Ideology	−0.220***	−0.200***	−0.213***	−0.204***
	(0.027)	(0.020)	(0.016)	(0.016)
Party Member	0.239***	−0.053		0.068
	(0.084)	(0.057)		(0.046)
Political Interest	0.050	0.027		0.046***
	(0.031)	(0.018)		(0.015)
Social Media Usage	−0.036	−0.053***		−0.053***
	(0.035)	(0.020)		(0.017)
Constant	4.043***	4.260***	4.314***	4.168***
	(0.228)	(0.151)	(0.045)	(0.126)
N	592	2,084	2,738	2,676
Adjusted R^2	0.146	0.070	0.066	0.083

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

Table B3: Treatment Effects on Regime Support: Central Government

| | Regime Support: Central Government | | | |
	Study 1	Study 2	Combined	Combined
Treatment	0.213***	0.066*	0.103***	0.100***
	(0.068)	(0.038)	(0.033)	(0.033)
Female	−0.018	0.156***		0.129***
	(0.070)	(0.039)		(0.034)
Education	0.047	−0.005		0.016
	(0.045)	(0.029)		(0.025)
Age Group	0.007	−0.044***		−0.043***
	(0.025)	(0.012)		(0.011)
Income	0.023	0.060***		0.045**
	(0.040)	(0.021)		(0.018)
Ideology	−0.198***	−0.239***	−0.228***	−0.217***
	(0.027)	(0.020)	(0.016)	(0.016)
Party Member	0.161*	−0.087		0.011
	(0.082)	(0.057)		(0.046)
Political Interest	0.030	0.063***		0.074***
	(0.030)	(0.018)		(0.015)
Social Media Usage	−0.030	0.024		0.0004
	(0.034)	(0.020)		(0.017)
Constant	4.178***	4.005***	4.399***	4.010***
	(0.224)	(0.150)	(0.045)	(0.126)
N	591	2,076	2,730	2,667
Adjusted R^2	0.121	0.094	0.074	0.093

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

Table B4: Treatment Effects on Regime Support: Local Government

| | Regime Support: Local Government | | | |
	Study 1	Study 2	Combined	Combined
Treatment	0.248***	0.085**	0.120***	0.121***
	(0.077)	(0.040)	(0.035)	(0.035)
Female	0.025	0.101**		0.091**
	(0.078)	(0.041)		(0.036)
Education	0.006	0.003		0.008
	(0.051)	(0.031)		(0.026)
Age Group	0.027	−0.021		−0.023**
	(0.028)	(0.013)		(0.011)
Income	−0.047	0.111***		0.079***
	(0.045)	(0.022)		(0.019)
Ideology	−0.256***	−0.221***	−0.243***	−0.232***
	(0.030)	(0.021)	(0.017)	(0.017)
Party Member	0.196**	−0.105*		−0.002
	(0.093)	(0.060)		(0.049)
Political Interest	0.025	0.033*		0.037**
	(0.034)	(0.019)		(0.016)
Social Media Usage	−0.024	−0.013		−0.018
	(0.038)	(0.021)		(0.018)
Constant	4.313***	3.819***	4.287***	3.950***
	(0.252)	(0.159)	(0.048)	(0.134)
N	580	2,079	2,721	2,659
Adjusted R^2	0.134	0.080	0.075	0.085

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

Table B5: Treatment Effects on Willingness to Protest

| | Willingness to Protest | | | |
	Study 1	Study 2	Combined	Combined
Treatment	−0.293***	−0.082	−0.130***	−0.124***
	(0.103)	(0.051)	(0.046)	(0.046)
Female	−0.124	−0.099*		−0.107**
	(0.105)	(0.053)		(0.047)
Education	0.012	0.055		0.015
	(0.069)	(0.040)		(0.034)
Age Group	0.067*	−0.014		−0.002
	(0.038)	(0.016)		(0.015)
Income	−0.033	0.093***		0.095***
	(0.060)	(0.028)		(0.025)
Ideology	−0.059	−0.130***	−0.136***	−0.129***
	(0.040)	(0.027)	(0.022)	(0.022)
Party Member	0.409***	−0.128*		−0.056
	(0.124)	(0.077)		(0.064)
Political Interest	−0.099**	−0.012		−0.046**
	(0.046)	(0.024)		(0.021)
Social Media Usage	−0.035	−0.077***		−0.054**
	(0.051)	(0.027)		(0.024)
Constant	3.325***	3.414***	3.430***	3.470***
	(0.338)	(0.203)	(0.062)	(0.175)
N	590	2,080	2,734	2,670
Adjusted R^2	0.030	0.034	0.016	0.028

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

B.2 Heterogeneous Treatment Effect

Figure B1 shows the heterogenous treatment effects among different demographic subgroups. As shown in the figure, treatment effects are weaker among respondents with lower education. This might indicate that (1) lower educated respondents are less able to pick up the treatment or (2) they are less susceptible to normalization. In the meantime, the confidence intervals of lower educated respondents are wider, suggesting that the weaker treatment effect might be due to insufficient sample size.

Figure B1: Heterogeneous Treatment Effects on Outcome Variables (Two Studies Combined)
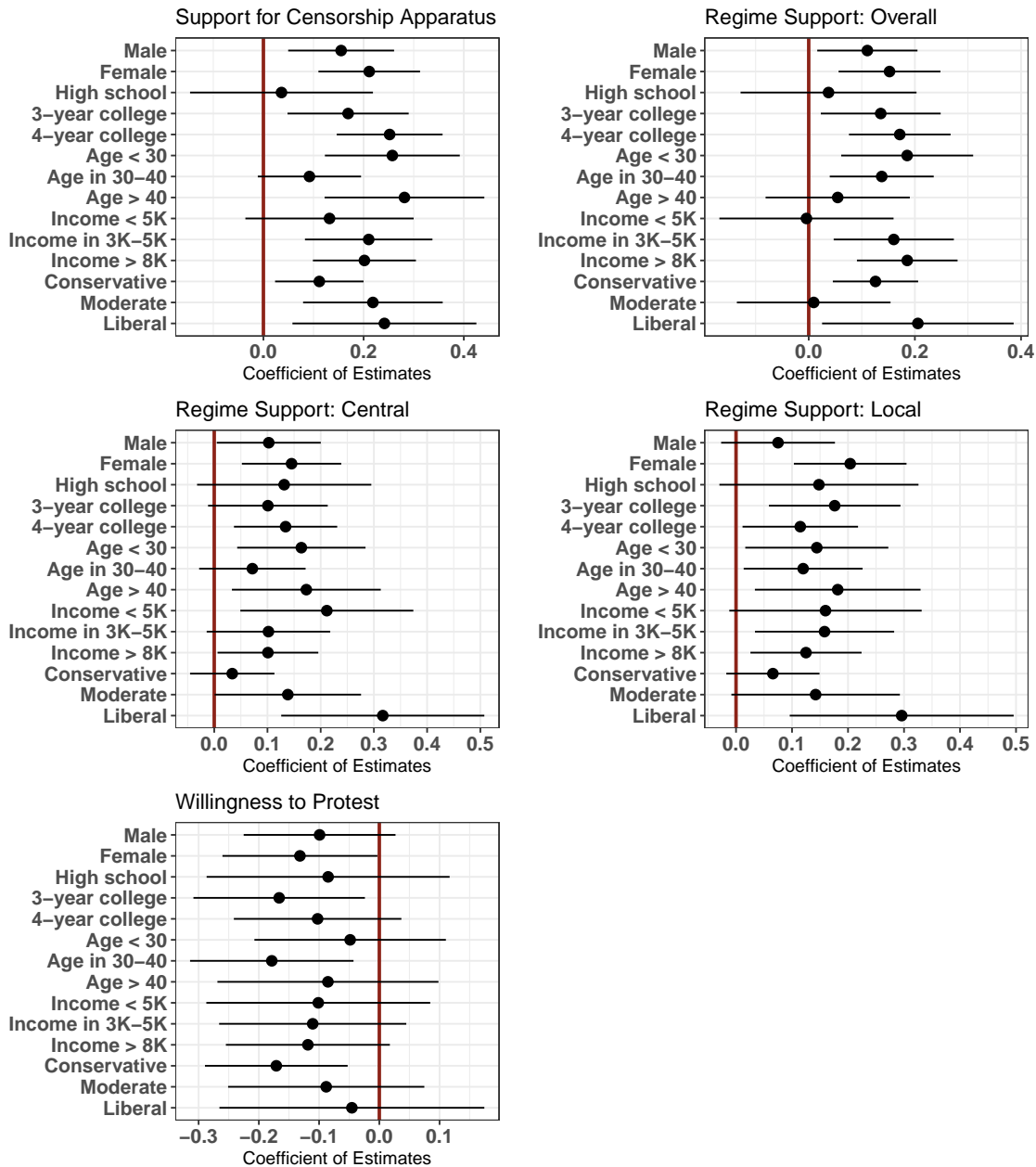
B.3 Study 2: Blank Control Group

As mentioned in the main paper, in addition to the control and treatment groups, I also include a blank control group in study 2, where respondents also read the same ten snippets but without any censorship labels. The purpose of the blank control group is to provide a benchmark for the other two groups. Comparing the blank control (without censorship) and the control group (with censorship of political content) also reaffirms the backlash effect of political censorship. In other words, I expect individuals exposed to censorship of political content will display lower support for both the censorship apparatus and the regime, compared with individuals who are not exposed to censorship at all.

As shown in Figure B2, censorship of political content did significantly decrease support for both the censorship apparatus and the regime. Moreover, additional censorship of non-political content brings the level of support back to the blank control level.

Figure B2: Group Means of the Dependent Variables (Blank Control Group Included)
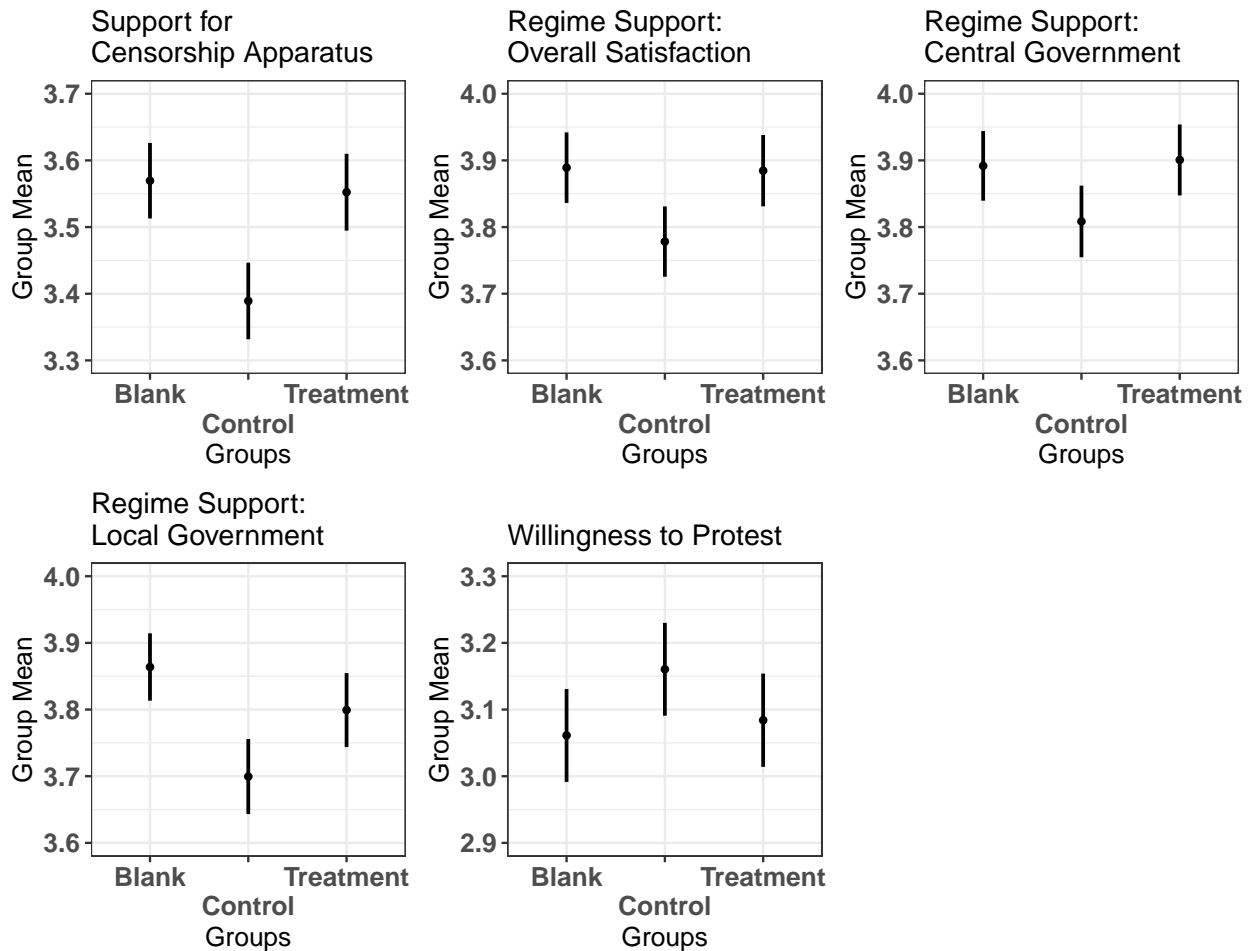
Note: All three outcome items are measured on a five-point Likert scale.

Appendix C   Experiments: Experiment Articles

As explained in the main paper, in both experiments, to expose participants to censorship, I asked respondents to read ten snippets of WeChat articles, presented one at a time with only the title and the first few lines. 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. Table C1 and Table C2 report the titles of articles used in both studies. For the first experiment, I selected snippets from the WeChatScope dataset used in the observational study. For the second experiment, I selected snippets from another website recording Chinese censorship. The selection process was systematic.

Table C1: Treatment Articles for Study 1 (Order Randomized)

#	Political Content	Control Group	Treatment Group	Title
1	No		Censorship Label	The banks are crying, one trick to help you earn 23 times higher profit by demand deposit.
2	No		Censorship Label	How much do men care about your face?
3	No		Censorship Label	The King of traditional medicine Sun Simiao lived 142 years. Before he died, he told his pupils: Be sure to destroy this prescription!
4	No			Ten questions about Huawei's former employees being sued for extortion.
5	No			Say goodbye to the stressful status quo. How can we relax under the pressure of work?
6	No			Thaksin and Yingluck returned to Meizhou to worship their ancestors.
7	Yes	Censorship Label	Censorship Label	Just now, the Pingshan Jasic labor strike has won an initial victory!
8	Yes	Censorship Label	Censorship Label	Jiangxi's "Funeral Reform" must not smash the coffin and hurt people's hearts
9	Yes	Censorship Label	Censorship Label	President Hu and Premier Wen are getting old. What happened in their decade?
10	Yes			After the tax reform, has your income decreased?

Table C2: Treatment Articles for Study 2 (Order Randomized)

#	Political Content	Control Group	Treatment Group	Title
1	No		Censorship Label	Please! Shut up! Don't Like Them Anymore!
2	No		Censorship Label	If you feel pain here, maybe problems with the meridian. Try this herb!
3	No		Censorship Label	Full-time housewife for 20 years and only get 50K for divorce. Brutal truth about marriage.
4	No			Please don't over-interpret Lu Daosen's suicide.
5	No			Wu Zhihong: Be careful of people with too much positive energy.
6	No			Why "Akita beauties" are not happy?
7	Yes	Censorship Label	Censorship Label	24 hours later, is the Xuzhou chained women in black doing okay?
8	Yes	Censorship Label	Censorship Label	Zhang Weiying: Democracy is a commitment.
9	Yes	Censorship Label	Censorship Label	White elephant projects are not government accomplishment.
10	Yes			Chen Jizhi said sorry after convicted, the same as Hu Xijin.

Appendix D Text Analysis: Categorization of Censored Articles

D.1 Categories and Coding Process

In total, I keep track of nine different topic categories. In addition to three highly political categories: (1) collective action, (2) government criticism, and (3) other government-related articles, I also include six moderately political and non-political categories: (1) business, (2) foreign events, (3) entertainment & sexuality, (4) advertisement, (5) local cultures & traditions, and (6) others. The practice to distinguish non-political content from political content is consistent with recent research on authoritarian censorship (Esberg 2020).

The categorization process and coding rubrics mainly follow Miller (2018), because Miller (2018) provides the most detailed, reliable, and up-to-date categorization of censored content in China. In particular, the definition of collective action, business, and entertainment & sexuality is the same as Miller (2018). The definition of government-related content combines the definitions of seven different categories in Miller (2018): *government*, *corruption*, *sensitive anniversary*, *recurring political event*, *regular political event*, *nationalism*, and *HK/Macau/Taiwan*. By using a broader definition of government-related content, I aim to establish the upper bound on the proportion of political content and avoid underestimating government-related content. The difference between government criticism and other government-related articles also follows the definition of government criticism in Miller (2018). Any government-related content that does not meet the definition of government criticism is categorized as other government-related articles.

The last four categories are created by myself due to the incompleteness of Miller's coding rules to my data. They are all self-explanatory. Importantly, none of these categories include politically salient events or issues. The coding rubrics for non-political categories explicitly exclude content related to the Chinese government. For example, the business category excludes government economic policies, state-own enterprises, and any mention of government institutions; the foreign events category requires the article to have no direct reference to China. The last category is the residual category which includes all articles that do not fit into the definitions of the other groups.

One important difference from Miller (2018) is that the nine categories are mutually exclusive. A similar strategy is employed by King, Pan, and Roberts (2013). Having mutually exclusive categories simplifies the categorization process as well as the interpretation of the results. In practice, the nine categories are coded sequentially with political categories coded first. Specifically, an article will first be considered if it belongs to the collective

action category. If yes, then the categorization process ends. If not, the article will then be considered if it belongs to the government criticism category and so on. If an article does not fit into the definitions of the first eight categories, it will be put into the last residual category. The coding process ensures that the analysis will not underestimate collective action and government criticism.

D.2 Inter-Coder Reliability

Two coders coded the 12,500 articles and posts (2,500 from WeChatScope and 5,000 from FreeWeChat and WeiboScope respectively) in the training set independently. To code the training set, they both analyze the titles, the authors, and the content of the articles according to the coding rubric. Both coders are native Chinese graduate students in political science. Table D1 shows that their results are generally consistent in terms of the proportion of each topic category. The greatest disagreement between the two coders is whether an article belongs to Government Criticism (CRI) or other government-related articles (GOV), which is not the main focus of this paper.

Table D1 shows the details of the two coders' coding. The accuracy rate between the two coders is 82.28% when considering specific topic categories. When identifying whether an article is political or non-political, the two coders agree on 93.28% of the cases. The macro F1 is 0.82 and the Cohen's κ between the two coders is 0.80, higher than the commonly applied criteria of 0.70 for inter-coder reliability tests. In cases where the two coders disagreed, the author acted as an arbitrator to settle the dispute.

Table D1: Inter-Coder Reliability

	ADS	BET	COL	CRI	ESX	FOR	GOV	LCT	OTH	Macro
Precision	0.91	0.89	0.86	0.95	0.91	0.92	0.65	0.80	0.58	0.83
Recall	0.86	0.89	0.90	0.70	0.89	0.78	0.87	0.77	0.78	0.83
F1	0.88	0.89	0.88	0.81	0.90	0.84	0.74	0.79	0.67	0.82

Note: ADS: Advertisement. BET: Business. COL: Collective Action. CRI: Government Criticism. ESX: Entertainment & Sexuality. FOR: Foreign Events. GOV: Government (Others). LCT: Local Cultures & Traditions. OTH: Others.

D.3 Example Articles

1. Collective Actions (COL)

   (a) 写入国务院文件的要求，群众拉横幅抗议！舆论哗然后教育部回应了
   The request written in the documents of the State Council; the masses raised banners to protest! Public opinion was upset and the Ministry of Education responded.

   (b) 刚刚，坪山佳士工人斗争取得初步胜利！
   Just now, the Pingshan Jiashi workers' struggle has won an initial victory!

   (c) 香港暴徒冲击手册曝光！密谋7·21漆弹袭警
   Hong Kong thugs' manual exposed! Plotting 7.21 Paintball Attack on Police.

   (d) 澳媒：香港暴力抗议者统治街头
   Australian media: Hong Kong violent protesters rule the streets.

   (e) 山东上访村民引爆炸药,与镇政府大院同归于尽你怎么看？欢迎转发！
   The petitioning villagers in Shandong Province detonated explosives, attempting to employ suicide bomb strategy at the township government. What do you think? Welcome to repost!

2. Government Criticism (CRI)

   (a) 武汉封城第43天，我问了方方一些日记之外的批评问题
   On the 43rd day of Wuhan's lockdown, I asked Fang Fang some critical questions outside of her diary.

   (b) 河南又火了，官员组团贪腐
   Henan province is under the spotlight again: collective corruption of the officials.

   (c) 征二胎税？！历史将永远记住这两个递刀的人！
   Second-child tax? ! History will always remember these two complicit!

   (d) 任志强国务院开炮，大胆发言
   Ren Zhiqiang blasts the State Council.

   (e) 41人任命被撤销，违规提拔后，被退回官员怎么办？
   41 appointments were revoked. What should we do with illegally promoted officials?

3. Government Non-Criticism (GOV)

   (a) 中国四中全会期待改革再出发，应对大变局政策需落地
   The Fourth Plenary Session of China's Central Committee looks forward to reforms, and policies to respond to major changes need to be implemented.

   (b) 一文让你看懂中国特色社会主义制度的优越性！
   This article allows you to understand the superiority of the socialist system with Chinese characteristics!

(c) 我国动真格了！直接取消德国千亿合同，令其肠子悔青，疾呼：有话好好说！
Our country is showing strength! Cancelling the 100 billion contract with Germany. The Germans are in regret!

(d) 台湾政坛大乱斗，大陆随时可能解放台湾！
The political chaos in Taiwan. Mainland may liberate Taiwan anytime!

(e) 好消息，中国！即将实现打电话不用钱，上网不用流量！
Good news, China! It will soon be realized that cellphone and Internet data are free!

4. Business (BET)

(a) 重磅!碧桂园再爆危机涉中信诚信窝案20亿利益输送或解地产界陈年谜案
Big News! Another crisis of the Country Garden Group, involving the 2 billion CITIC case.

(b) 谷歌或推出中国版搜索引擎茅台，百度股价闻声大跌近8%！
Google may launch the Chinese version of the search engine: Maotai and Baidu's stock price plummeted by nearly 8%!

(c) 陈浩武：杨小凯经济学思想的时代意义（下）
Chen Haowu: The Significance of Yang Xiaokai's Economic Thoughts (Part 2)

(d) 这次章泽天输给了董明珠财经要参
This time, Dong Mingzhu beats Zhang Zetian — Finance insider

(e) 滴滴总裁柳青发了封道歉信，湖畔大学学员：加油
Liu Qing, CEO of Didi, issued a letter of apology.

5. Entertainment & Sexuality (ESX)

(a) 男人究竟有多在意你的脸?
How much do men care about your face?

(b) 香港绝色女艳星：喂春药令李小龙命丧闺床，死后大拍激情片…
Hong Kong's stunning female porn star: Bruce Lee was killed by feeding aphrodisiac, making a passionate movie after his death...

(c) 台湾女教授深入讲解男女关系，太精彩了！不浪费一秒钟从头笑到尾
Taiwanese female professors explain in depth the relationship between men and women: it is wonderful! Don't waste a second laughing from start to finish

(d) 贾乃亮助理点赞李小璐不要脸? 刘亦菲曾被劈腿? 范冰冰姐弟关系揭秘?
Jia Nailiang liked Li Xiaolu as shameless? Liu Yifei was cheated? Fan Bingbing's relationship between sister and brother revealed?

(e) 48岁的闫妮惊艳春晚：瘦了30斤之后，气质简直比20岁的女儿还好！
48-year-old Yan Ni was amazing at the Spring Festival Gala: After losing 30 pounds, her temperament is better than her 20-year-old daughter!

6. Advertisement (ADS)

(a) 3斤枸杞不如1两它，最强天然青霉素，每天喝一杯，抵抗力强10倍！

3 pounds of wolfberry is not as good as 1 tael of the natural penicillin. Drink a cup per day to strengthen your immunity 10 times!

(b) 健康—经期来了可别再喝红糖水了，这样做排出体内十年垃圾

Health—Don't drink brown sugar water anymore when your menstrual period is coming. Do this to get rid of the garbage in your body for ten years

(c) 【强烈推荐】这东西补血养颜，女人要多吃，40岁像20岁！

Strongly recommended! This food nourishes blood and beauty. Women should eat more. Makes you look like 20 years old when you are 40!

(d) 【推荐】原来明星大V们喜欢关注这些公众号

Recommended! It turns out that celebrities like to pay attention to these public accounts

(e) 第一次创业，我建议你选择小程序代理W

For the first time starting a business, I suggest you choose a mini program agent

7. Local Cultures & Traditions (LCT)

(a) 故宫的规矩

Rules of the Forbidden City.

(b) 辣眼睛！幼儿园开学典礼竟大跳钢管舞！园长被开除...

OMG! Pole dance in kindergarten! The principal was fired...

(c) 难怪民国出大师，民国老课本美到心醉。现在的孩子都该看看

No wonder there were many masters during the Republic era. These old textbooks from the Republic era are so beautiful. Children should read them now.

(d) 6 Breakfast Items You Must Try in Beijing

(e) 许倬云：中国文化让我伤心的地方

Xu Zhuoyun: What makes me sad about Chinese culture.

8. Foreign Events (FOR)

(a) 特朗普遭美国众议院弹劾，真能弹掉吗？

Trump was impeached by the US House of Representatives, but can he really be impeached?

(b) 德国的吸星大法：元首当年用铁和血没做到的事情，被你们用欧元做到了！

The German method of attracting money: What Hitler failed to do with iron and blood, was done by the Germans with euros!

(c) 苏莱曼尼算死于暗杀吗？

Does the killing of Soleimani count as assassination?

(d) 日本是如何对待黑人与穆斯林的

How Japan treats the Black and Muslins.

(e) 日本海底核试验败露!地震后的奇异景象已证实?安倍这次该如何收场？

Japan's underwater nuclear test revealed! The bizarre scene after the earthquake has been confirmed? How should Prime Minister Abe react this time?

9. Others (OTH)

   (a) 她不是一个人在战斗！
       She is not fighting alone.

   (b) 师恩·难忘
       Teacher's kindness, unforgettable

   (c) 国人为何热衷于炫耀?
       Why are Chinese people keen on showing off?

   (d) 最后，只有东北人没有被导游忽悠到
       In the end, only the Northeast people were not fooled by the tour guide

   (e) 马基雅维里——恶魔还是巨人?
       Machiavelli — Demon or Giant?

Appendix E Text Analysis: Models & Robustness

E.1 BERT Model Performance

To validate the performance of the fine-tuned pre-trained Chinese BERT with the Whole Word Masking model, I employed the human coders to further annotate a random sample of 5,000 machine-labeled data. I then calculate the precision, recall, and F1 score of the model performance. As shown in Table E1, the fine-tuned Chinese BERT model achieved a pretty high level of accuracy. the Macro precision, recall, and F1 are all around 0.94.

Table E1: Performance of the Fine-Tuned Chinese BERT Model

	ADS	BET	COL	CRI	ESX	FOR	GOV	LCT	OTH	Macro
Precision	0.96	0.95	0.97	0.86	0.96	0.98	0.94	0.93	0.93	0.9436
Recall	0.96	0.94	0.89	0.92	0.96	0.97	0.90	0.96	0.92	0.9344
F1	0.96	0.95	0.93	0.89	0.96	0.98	0.92	0.94	0.93	0.9386

Note: ADS: Advertisement. BET: Business. COL: Collective Action. CRI: Government Criticism. ESX: Entertainment & Sexuality. FOR: Foreign Events. GOV: Government (Others). LCT: Local Cultures & Traditions. OTH: Others.

E.2   Alternative Models

As a robustness check, I use multiple alternative models. First, I trained a multinomial logistic regression model with a ridge estimator. I chose penalized regression models were used because the number of predictors (text) is much larger than the number of observations. Since I do not wish to drop predictors in the regularization process, L2 ("ridge") penalty is preferable to L1 ("LASSO") penalty. The training model is specified as:

$$y_{ij} = \alpha_j + \mathbf{DFM}_i\beta_{\mathbf{j}} + \epsilon_{ij}$$

where y_{ij} is a binary variable that takes 1 if observation i belongs to topic category j and 0 otherwise. **DFM** is the document-feature matrix of the labeled data. **X** is a matrix of additional predictors. $\beta_{\mathbf{j}}$ is the matrix of ridge estimators for category j. Once the best matrices of ridge estimators, $\hat{\beta}_{\mathbf{j}}$, was found, I matched the unlabeled text corpus with the DFM of the labeled data. I then used the matched matrix and the best matrix of ridge estimators, $\hat{\beta}_{\mathbf{j}}$, to predict the unlabeled data.

Before the text analysis, all punctuation and stop words are removed and the Chinese text is segmented into individual tokens. Then, the segmented text was converted into a document-feature matrix. Words that appear less than 4 times were removed from the document-feature matrix.

I also train two additional sets of alternative models. They are linear regression models with ridge estimators and binomial logistic regression models with ridge estimators. The procedures are mostly the same as the multinomial ridge models. However, each topic category is predicted independently. The topic category with the highest predicted value is chosen to be the predicted topic category.

Table E2 shows that predictions are generally consistent across the three models. All three models predict highly politically threatening content to account for around 33-40% of all censored articles. Confirming the theoretical expectation that non-political content accounts for the majority of all censored content.

Table E2: Predicted Proportion of Censored Articles by Topic Category – Alternative Models

	Multinomial	Linear	Binomial
Collective Action	0.67%	1.13%	1.01%
Government Criticism	27.94%	21.59%	27.69%
Other Government-related	10.73%	11.07%	9.76%
Business	14.43%	16.36%	15.15%
Foreign Events	3.38%	3.77%	4.08%
Entertainment & Sexuality	20.32%	26.21%	20.66%
Advertisement	7.40%	8.00%	7.98%
Cultures & Traditions	12.28%	11.21%	12.27%
Others	2.86%	0.67%	1.40%

Table E3: Top Keywords of Each Topic Category

Category	Keywords
Collective Action	暴徒 Violent Protesters 维权 Rightful Resistance 香港 Hong Kong 工会 Labor Union
Government Criticism	落马 Removed from Office 违规 Violating Laws 主任 Official 任志强 Ren Zhiqiang 纪委 Commission for Discipline Inspection
Other Government-related	习近平 Xi Jinping 市长 Mayor 共青团中央 CCP Youth League 中方 Chinese side 改革开放 Reform and Openning Up
Business	谷歌 Google 马云 Jack Ma 股市 Stock Market 研发 Innovation 上市 IPO 京东 JD.com 资金 Capital 网易 NetEase Inc
Foreign Events	朝鲜 North Korea 全世界 The World 日本 Japan 特朗普 Donald Trump 土耳其 Turkey
Entertainment & Sexuality	老婆 Wife 单身 Single 吴秀波 Wu Xiubo 爱情 Love 娱乐圈 Entertainment Circle
Advertisement	优质 Good Quality 轻松 Easy 睡眠 Sleeping 肠胃 Stomach 蜂蜜 Honey
Local Cultures & Traditions	景区 Tourist Attractions 道德 Morality 公交 Public Transportation 中医 Chinese Medicine

Note: Keywords are extracted from linear ridge models and binomial ridge models. Keywords are among the top predictors of each topic category.

To validate the multinomial ridge model prediction, I conduct cross-validation using the WeChatScope data and the multinomial model predictions. I randomly divide the 15,872 articles into three groups. Each time, I use two groups as the training set to predict the remaining group. Table E4 reports the cross-validation results. In general, the cross-validation results suggest that the model performance is not as good as the state-of-the-art BERT model. Nevertheless, the macro F1 is still above 0.7, indicating a moderately good job in classification.

Table E4: Cross-Validation for Multinomial Ridge Models

	ADS	BET	COL	CRI	ESX	FOR	GOV	LCT	OTH	Macro
Precision	0.84	0.79	0.93	0.72	0.81	0.75	0.76	0.76	0.72	0.79
Recall	0.72	0.76	0.41	0.86	0.82	0.51	0.69	0.73	0.41	0.66
F1	0.78	0.76	0.57	0.78	0.81	0.61	0.73	0.74	0.52	0.70