Normalization of Censorship: Evidence from China Online Appendices

Tony Zirui Yang

February 12, 2024

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.

Both survey experiments were pre-registered prior to the implementation of the surveys. The anonymized pre-analysis plan of study 1 can be found here:
`https://osf.io/73ej5/files/osfstorage/63f8e804bbc5e502ccf80265`.
The anonymized pre-analysis plan of study 2 can be found here:
`https://osf.io/4pg8f/?view_only=51d529726e464cd7a22ff5565fbd2fee`.

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 can 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 that 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 would 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%	54.5%	46.2%
	Central	19.6%	21.5%	22.1%
	West	21.6%	17.5%	23.3%
	Northeast	7.8%	5.9%	8.4%
Gender	Female	49.7%	49.7%	48.1%
	Male	49.7%	49.8%	51.9%
Education	≤ Junior high	3.6%	3.7%	56.1%
	Senior high	12.6%	16.4%	23.8%
	3-year college	25.2%	36.3%	10.5%
	≥ 4-year college	58.5%	43.3%	9.7%
Age	≤ 19	6.5%	2.6%	23.2%
	20-29	31.4%	27.3%	21.5%
	30-39	45.1%	46.0%	20.8%
	40-49	14.9%	15.9%	17.6%
	≥ 50	2.1%	8.2%	16.9%
Income	≤ 3000	7.8%	6.3%	51.0%
	3000-5000	13.9%	13.8%	21.5%
	5000-8000	38.2%	32.4%	14.3%
	≥ 8000	38.9%	47.0%	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	0.248	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	2.345	.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	.54	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

7

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)
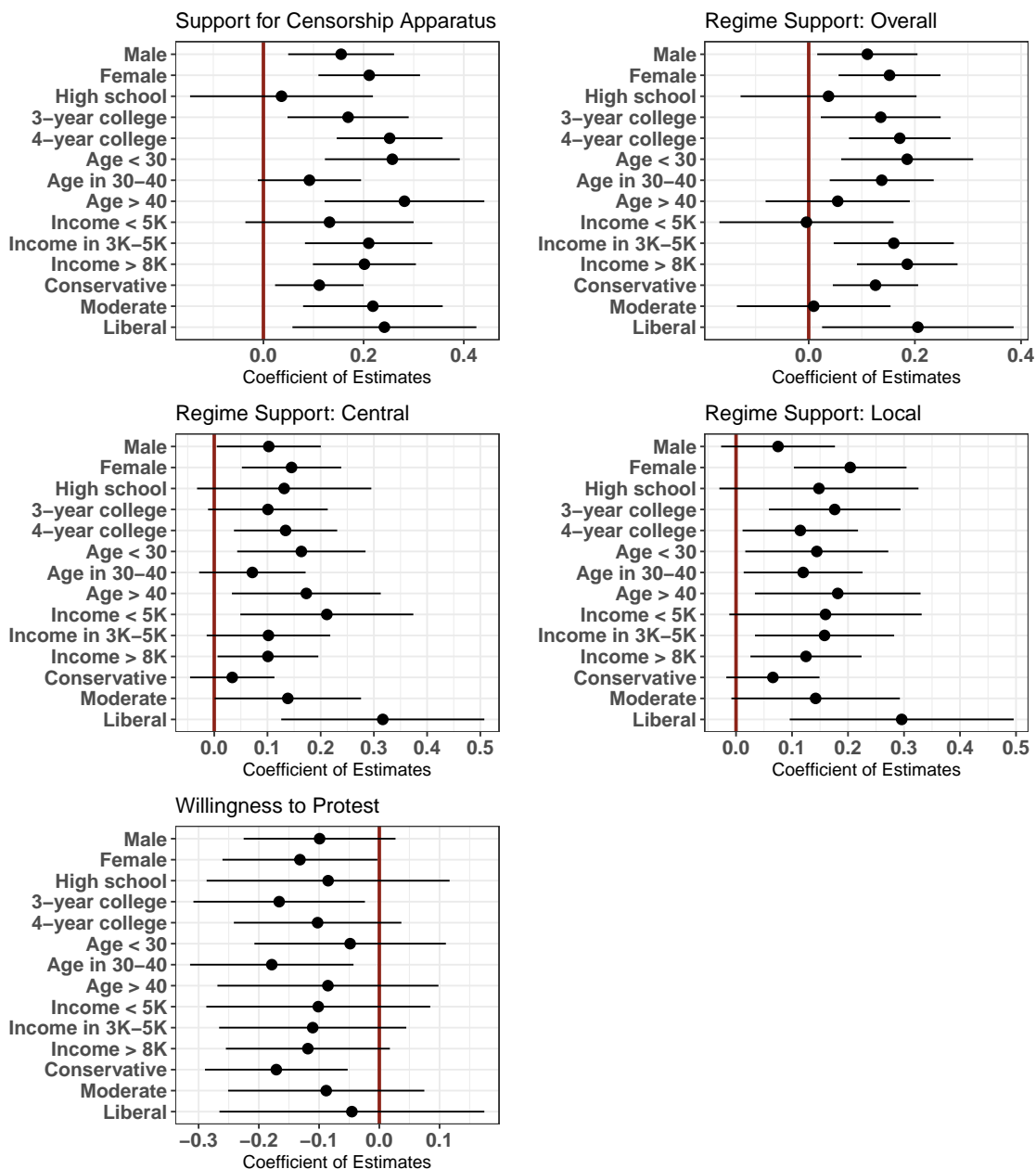
B.3 Multiple Hypotheses Testing Correction

I used the Benjamini-Hochberg (BH) p-value correction method to account for multiple comparisons. As shown in Table B6, all existing significant results survive the BH correction and are still significant at the conventional level.

Table B6: Multiple Hypotheses Testing Correction (Benjamini-Hochberg)

	Treatment Effects		
Dependent Variable	*Study 1*	*Study 2*	*Combined*
Support for Censorship Apparatus	0.264	0.163	0.185
p-value	[0.00144]	[0.00010]	[0.00000]
adjusted p-value	[0.00270]	[0.00043]	[0.00001]
Overall Satisfaction of China	0.229	0.106	0.133
p-value	[0.00183]	[0.00604]	[0.00011]
adjusted p-value	[0.00306]	[0.00906]	[0.00043]
Assessment of Central Government	0.236	0.092	0.124
p-value	[0.00097]	[0.01821]	[0.00031]
adjusted p-value	[0.00208]	[0.01951]	[0.00092]
Assessment of Local Government	0.288	0.100	0.141
p-value	[0.00037]	[0.01457]	[0.00012]
adjusted p-value	[0.00092]	[0.01681]	[0.00043]
Willingness to Protest	-0.268	-0.076	-0.118
p-value	[0.00852]	[0.13341]	[0.01021]
adjusted p-value	[0.01162]	[0.13341]	[0.01276]

B.4 Implicit Support for Censorship

In addition to the additional analyses in Section 5.5 of the main paper, to further alleviate the concerns about preference falsification, I also use a list experiment to measure implicit support for censorship in Study 2. The list experiment uses the exact same wording as the censorship support question. Figure B2 the results for implicit and explicit support for the censorship apparatus in Study 2, where explicit support is the proportion of respondents who chose somewhat or strongly support, in the explicit question. If we only examine the point estimates, in the control group, around 46% of the respondents exhibit implicit support for censorship, whereas in the treatment group, this figure rises to 51%. Comparing these results to the explicit support for censorship at 53% for the control group and 58% for the treatment group, two key observations can be made. First, the overall level of preference falsification is at most 7 percentage points. This suggests that, on the whole, preference falsification may not be a pervasive issue. Second, we still detect a similar increase in implicit support when moving from the control group to the treatment group, indicating that the level of preference falsification is not higher in the treatment group. However, the biggest problem with the current list experiment is that their estimates are imprecise. Specifically, the 95% confidence interval for implicit support in the control group is [37%, 57%], and [41%, 62%] for the treatment group. These intervals encompass over 20 percentage points, making it challenging to find any statistically significant treatment effects using implicit measures.

Figure B2: Implicit Support for Censorship

13
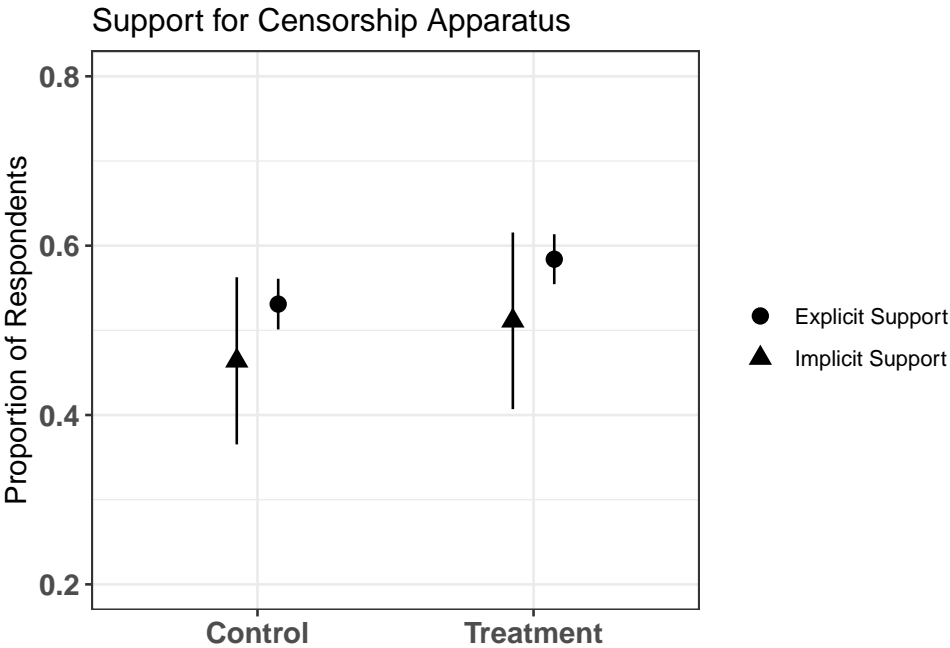
Appendix C Experiments: Experiment Articles

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, (4) advertisement, (5) cultures, and (6) others. The practice of distinguishing 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 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 were 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-owned 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 82.5% when considering specific topic categories. When identifying whether an article is political or non-political, the two coders agree on 92.97% 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	0.92	0.87	0.88	0.94	0.89	0.97	0.68	0.84	0.50	0.83
Recall	0.86	0.90	0.91	0.71	0.89	0.78	0.87	0.79	0.76	0.83
F1	0.89	0.88	0.90	0.81	0.89	0.86	0.77	0.81	0.60	0.82

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

D.3 Content within Each Topic Categories

To better understand what kinds of content are being censored in each specific topic category, I run simple Structural Topic Models (STM) within each of the topic categories. I then manually identified the 10 most common topics among the STM results, along with their associated keywords.

Collective Action

- Topic 1: Hong Kong protest movement in 2019

    - Keywords: HK, violent protesters, riot, protest, police, terrorists, looting

- Topic 2: Other protests in Hong Kong

    - Keywords: HK, Pan Democrats, July 1st, gather, independence, LegCo

- Topic 3: Labor strikes

    - Keywords: labor union, workers, factory, Shenzhen, employee, stop production

- Topic 4: Historical revolutions in China

    - Keywords: revolution, Opium War, Boxer Rebellion, Red Guard, the West

- Topic 5: Uyghur unrest

    - Keywords: Xinjiang, Uyghur, sovereignty, riot, public security

- Topic 6: Weiquan movement and petitioning

    - Keywords: rightful resistance, petitions

- Topic 7: COVID-related collective actions

    - Keywords: COVID, quarantine, testing, Fangcang hospital

- Topic 8: Picking quarrels and provoking trouble

    - Keywords: bully, violence, beer bottle, while shirt, police, incident

- Topic 9: Foreign involvement in collective actions

    - Keywords: spy, CIA, United States, US Congress, Trump

- Topic 10: Bank-related collective actions

– Keywords: deposit, bank, gather, chanting, migrant workers

Government Criticsm

- Topic 1: COVID-19 pandemic (initial outbreak)

  – Keywords: Li Wenliang, lockdown, Wuhan, Fangcang hospital, Fang Fang

- Topic 2: COVID-19 pandemic (criticism of zero-COVID policy)

  – Keywords: vaccine, zero-COVID, Omicron, Pfizer, case, positive case

- Topic 3: Corruption

  – Keywords: bribery, violation of discipline, take bribes, discipline inspection

- Topic 4: Criticism of the one-child policy

  – Keywords: one-child policy, aging population, birth rate, population growth

- Topic 5: Criticism of the Xiong'An (planned new capital) policy

  – Keywords: Xiong'An, Xiong County, new district, demolition, flooding, villagers

- Topic 6: Criticism of foreign policies

  – Keywords: wolf warriors, little pinky, Hu Xijin,

- Topic 7: Feminism

  – Keywords: women's rights, patriarchy

- Topic 8: Democratic values

  – Keywords: freedom, human rights, liberalism, property rights

- Topic 9: Criticism of the police and censorship

  – Keywords: questioning, the police, netizens, demand, media, investigation, misinformation, censorship, rumor

- Topic 10: Criticism of economic policies

  – Keywords: labor force, employment, pension, fiscal policy, manufacturing

Other Government-Related

- Topic 1: Party leaders

    – Keywords: Xi Jinping, General Secretary, Hu Jintao, Jiang Zemin, Zhu Rongji

- Topic 2: Communist Party

    – Keywords: Party organization, secretary, comrade, appoint, decide, economic development, work, meetings, reform

- Topic 3: Court

    – Keywords: Supreme People's Court, Appeal, Defendant, Laywer, Imprisonment

- Topic 4: Taiwan

    – Keywords: Taiwan, brainwash, defame, anti-China, nation, reunification

- Topic 5: Xinjiang

    – Keywords: Xinjiang, autonomous region, Hetian, safeguard, Terrorists

- Topic 6: Ideology

    – Keywords: Marxism, Mao Zedong, Dong Xiaoping, socialism, capitalism

- Topic 7: History of the Communist Party

    – Keywords: Yan'an, Chairman Mao, Lin Biao, Kuomintang, WWII, history

- Topic 8: Military

    – Keywords: PLA, tank, helicopter, navy, fighter jets, air force

- Topic 9: COVID-19 Pandemic

    – Keywords: vaccine, virus, WHO, immune, mutation, case

- Topic 10: US-China Trade War

    – Keywords: Huawei, trade war, Trump, sanction, chips, RMB

Business

- Topic 1: Investment Tips (Stock Market)

– Keywords: Index Fund, Stock price, Bond, IPO, Long, Short, Buy-in, Sell-out

- Topic 2: Investment Tips (Real Estate)

    – Keywords: Housing price, Second-hand house, Buying house, School district

- Topic 3: Investment Tips (Crypto Currency)

    – Keywords: Cryto, Dogecoin, Encrypted, Block-chain, Mining, Elon Mask

- Topic 4: Sector Analysis (Platform Companies)

    – Keywords: Meituan, Rider, Kuaishou, JD.com, Douyin, Pinduoduo, Tencent

- Topic 5: Sector Analysis (Food Industry)

    – Keywords: Ruixin Coffee, Starbucks, Nestle, Mooncakes, Brand, Ice cream, Yili

- Topic 6: Sector Analysis (Alcohol Industry)

    – Keywords: Maotai, Alcohol, Baijiu Liquor, Fermentation, Baijiu Aroma, Beer

- Topic 7: Sector Analysis (Electric Cars)

    – Keywords: Tesla, NIO, XPeng, New energy, Electric cars, Self-driving

- Topic 8: Sector Analysis (Education Industry)

    – Keywords: New Oriental, Yu Minhong, Education, Training, Extracurricular

- Topic 9: The Effect of COVID on Investment

    – Keywords: Vaccine, Virus, Coronavirus, Global, Economy

- Topic 10: Investment Tips (Trust & Equity)

    – Keywords: Trust, Billion Yuan, Shareholder, Bank

Foreign Events

- Topic 1: Domestic Politics of the United States

    – Keywords: Democrats, Republicans, midterm election, racism, BLM, conservative

- Topic 2: Domestic Politics of European Countries

– Keywords: Germany, Euro, Italy, France, Spain, European Union

- Topic 3: Domestic Politics of Russia & Russo-Ukrainian Conflict

  – Keywords: Russia, Putin, Crimea, Ukraine, Donbas, Kyiv

- Topic 4: Overseas Chinese Community Information (US)

  – Keywords: New York, Flushing, Brooklyn, Queens, Chinatown, Los Angeles, Southern California, house rental, restaurants

- Topic 5: Israel-Palestinian Conflict

  – Keywords: Israel, Hamas, Palestine, Syria, Gaza

- Topic 6: Other Events in the Middle East

  – Keywords: Iran, Iraq, Shia, Sunni, Ali Khamenei

- Topic 7: COVID-19 Pandemic in Foreign Countries

  – Keywords: mask, vaccine, infection, CDC, Delta, Omicron

- Topic 8: Overseas Chinese Community Information (Europe)

  – Keywords: Madrid, Catalonia, the Spanish-Chinese community

- Topic 9: Domestic Politics of India

  – Keywords: Modi, India, New Delhi, Nepal, Indian government

- Topic 10: Domestic Politics of Korea & Japan

  – Keywords: Yoshihide Suga, Shinzo Abe, Moon Jae-in, LDP

Entertainment

- Topic 1: Discussion of Movies

  – Keywords: Douban, movies, actors, actress, critics, movie festival, box office, Ashes of Time, Academy Awards

- Topic 2: Discussion of Entertainment Shows

- Keywords: reality shows, entertainment shows, Xiao Zhan, fans, stars, Sina Weibo, Guo Degang, New Year's Gala

- Topic 3: Tabloid Gossips

  - Keywords: gossip, paparazzi, fans, Zhao Wei, Guo Jingming, Xiao Yaxuan, Lin Zhixuan, Nicholas Tse

- Topic 4: Discussion of Classical Novels

  - Keywords: Dream of the Red Chamber, The Legend of the Condor Heroes, Wolf Totem, The Three-Body Problem

- Topic 5: Discussion of Beauty Standard

  - Keywords: whitening, loss of weight, skin, model

- Topic 6: Discussion of TV Series

  - Keywords: House of Cards, Prison Break, Breaking Bad, Spartacus

- Topic 7: Discussion of Documentaries

  - Keywords: documentaries, BBC, world, Renaissance, WWII, history

- Topic 8: Discussion of Music

  - Keywords: popular music, Wang Mingquan, Tie Xue Dan Xin

- Topic 9: Discussion of Relationships

  - Keywords: husband, wife, boyfriend, girlfriend, love, life, marriage

- Topic 10: Discussion of Talk Shows

  - Keywords: Liang Wendao, Gao Xiaosong, Behind the Headlines with Wen Tao

Advertisement

- Topic 1: Product Promotion: Food

  - Keywords: taste, crawfish, meat, chicken feet, eel, strawberry, corn, fruit

- Topic 2: Product Promotion: Courses & Training

– Keywords: textbook, vocabulary, multi-media, third grade, training, register

- Topic 3: Job Ads

    – Keywords: hiring, written examination, salary, earnings, position, qualifications

- Topic 4: Product Promotion: Household Product

    – Keywords: teapot, table, mask, sanitizer, toothpaste

- Topic 5: Product Promotion: Beauty Products

    – Keywords: sunscreen, lipstick, moisturizer, face mask, skin

- Topic 6: Product Promotion: Clothing

    – Keywords: material, pants, underwear, T-shirt, lightweight

- Topic 7: Product Promotion: Beverage

    – Keywords: tea, white tea, black tea, Pu'er tea

- Topic 8: Product Promotion: Medical Service

    – Keywords: eye hospital, gout, diabetes, nearsighted, orthopedics

- Topic 9: Product Promotion: Prescription Drugs

    – Keywords: ointment, bacteria, analgesics

- Topic 10: Product Promotion: Herbal Medicine

    – Keywords: Chinese medicine, cordyceps, goji berry, moisten the lungs, drink

Cultures

- Topic 1: Stories of Traditional Chinese Medicine

    – Keywords: master, apprentices, traditional medicine, secret recipe

- Topic 2: Stories of Chinese Poets

    – Keywords: Su Shi, Su Dongpo, Wang Anshi, Li Shangyin, Su Zhe, Tang Dynasty

- Topic 3: Feng Shui Stories

- Keywords: Feng Shui, secret, fortune, Chinese Zodiac, culture

- Topic 4: Stories of Fictional Figures in Chinese Literature

    - Keywords: Lin Daiyu, Jia Baoyu, Xue Baochai, Jin Ping Mei,

- Topic 5: Stories of Chinese Emperors

    - Keywords: Yongzheng, Kangxi, Qianlong, Liu Bang, Xiang Yu

- Topic 6: I Ching Stories

    - Keywords: I Ching, hexagram, wisdom, culture

- Topic 7: Bible Stories

    - Keywords: God, Jesus, Jehovah, Christ, Gospel

- Topic 8: Buddhist Stories

    - Keywords: Heart Sutra, Sarira, bodhi, incantation

- Topic 9: Taoist Stories

    - Keywords: Taoism, Lao Zi, Zhuang Zi

- Topic 10: Local Architecture

    - Keywords: architecture, photography, art, urban, history

Appendix E Text Analysis: Models & Robustness

E.1 Model Selection

To select the best classification model, I used the training data to test nine different machine-learning models. As shown in Table E1, the fine-tuned pre-trained Chinese BERT with the Whole Word Masking model is by far the best-performing model evaluated by out-sample five-fold cross-validation macro F1 score.

Table E1: Macro F1 Scores for Five-fold Cross-Validation

Model	Macro F1 Score
Fine-tuned Pre-Train Chinese BERT	0.7025
Logistic Regression (Ridge)	0.4628
Pattern Learning and Matching (PaLM)	0.4429
Extreme Gradient Boosting (XGBoost)	0.4363
Random Forest	0.4242
Ensemble Classifier (Voting)	0.4194
Decision Tree	0.4175
Neural Network	0.3558
Word2Vec Embedding	0.1763

E.2 BERT Model Performance

Based on the model selection results, I chose the fine-tuned pre-trained Chinese BERT with the Whole Word Masking model in the main analysis. The Chinese BERT model is a state-of-the-art deep learning model based on the Transformer architecture and pre-trained on a massive amount of Chinese text data. BERT learns contextualized representations of words by considering the entire sentence, capturing complex relationships between words. For text classification, the model takes the input text, tokenizes it, and passes it through its layers to produce embeddings. The embeddings are then used for topic classification tasks. The in-sample accuracy rate of the BERT model is 0.96 and the out-sample performance is presented in Table E2.

However, one concern arises regarding potential imbalances within the nine categories. To address this issue, I calculate category-specific weights to adjust for variations in category sizes. This approach allows for a more accurate assessment of the balanced performance of the BERT model. The results indicate a balanced precision of 0.71, a balanced recall of 0.73, and a balanced macro F1 score of 0.72.

Table E2: Out-sample Five-fold Cross-Validation

	ADS	BET	COL	CRI	ESX	FOR	GOV	LCT	OTH	Macro
Precision	0.73	0.74	0.70	0.52	0.77	0.89	0.68	0.65	0.70	0.71
Recall	0.81	0.72	0.75	0.50	0.76	0.87	0.70	0.64	0.70	0.72
F1	0.77	0.73	0.72	0.51	0.76	0.88	0.69	0.65	0.70	0.71

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

Furthermore, I combine the three highly political categories, as well as the remaining six non-political categories, transforming the classification task into a binary one. When determining whether an article falls into the highly political category or not, the BERT model excels with a balanced precision of 0.83, a balanced recall of 0.83, and a balanced macro F1 score of 0.83.

E.3 Logistic Regression Model with Ridge Estimator

As a robustness check, I use the second-best-performing model, the multinomial logistic regression model with a ridge estimator, to re-run the classification task using one of the three data sources. I chose penalized regression models 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, the L2 ("ridge") penalty is preferable to the 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. $\mathbf{DFM}$ is the document-feature matrix of the labeled data. $\mathbf{X}$ 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, $\hat{\beta}_{\mathbf{j}}$, were 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.

Table E3 shows that predictions are generally consistent with the main findings, with highly politically threatening content accounting for less than 40% of all censored articles. This confirms the theoretical expectation that moderate and non-political content accounts for the majority of all censored content.

Table E3: Predicted Proportion of Censored Articles by Topic Category – Alternative Models

General Category	Specific Category	Logistical Regression (Ridge)
Highly Political	Collective Action	0.71%
	Govt Criticism	26.71%
	Other Govt-related	10.91%
	Total	38.33%
Moderately Political	Business	14.48%
	Foreign	3.89%
	Total	18.37%
Non-Political	Entertainment	19.38%
	Advertisement	7.79%
	Culture	12.74%
	Others	3.38%
	Total	43.29%

Notes: Data Source: WeChatScope, 15,872 censored articles.