Participatory Censorship in Authoritarian Regimes Online Appendices

Tony Zirui Yang

October 28, 2024

Contents

Appendices

Appendix A Compliance with Ethical Principles of Human Subject Research

The two surveys conducted in this study followed all established principles of human subject research and were approved by the Institutional Review Board at Washington University in St. Louis. Although the IRB exempted this study from a formal consent form, I still included a consent page and information sheet at the beginning of the survey. All participants were informed about the purpose, the length, and the 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.

In the survey experiment (study 2), I showed respondents simulated social media posts that I adapted from Sina Weibo. Respondents are fully informed about how these posts are created from real posts, and therefore no deception is used.

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 this survey was 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' identity, 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.

Appendix B Study 1: Sample and Weighting

Table B1: Descriptive Statistics of the Original and Weighted Survey Sample (N=1,124)

Sociodemographic Variables		Original Survey Sample	Weighted Survey Sample	China Internet Census
Gender	Female	44.9%	45.2%	47.3%
	Male	54.8%	54.8%	52.7%
Location	Rural	30.7%	29.9%	28.2%
	Urban	67.9%	70.1%	71.8%
Region	East	25.2%	29.9%	31.1%
	South & Central	35.1%	29.5%	28.2%
	North & Northeast	26.1%	21.4%	22.2%
	West	12.7%	19.3%	18.5%
Age	≤ 19	8.6%	22.1%	21.6%
	20-29	40.0%	26.8%	26.8%
	30-39	32.5%	22.3%	23.5%
	≥ 40	17.4%	28.8%	28.1%
Education	$\leq$ High School	26.1%	77.4%	79.8%
	$\geq$ College	73.9%	22.6%	20.2%

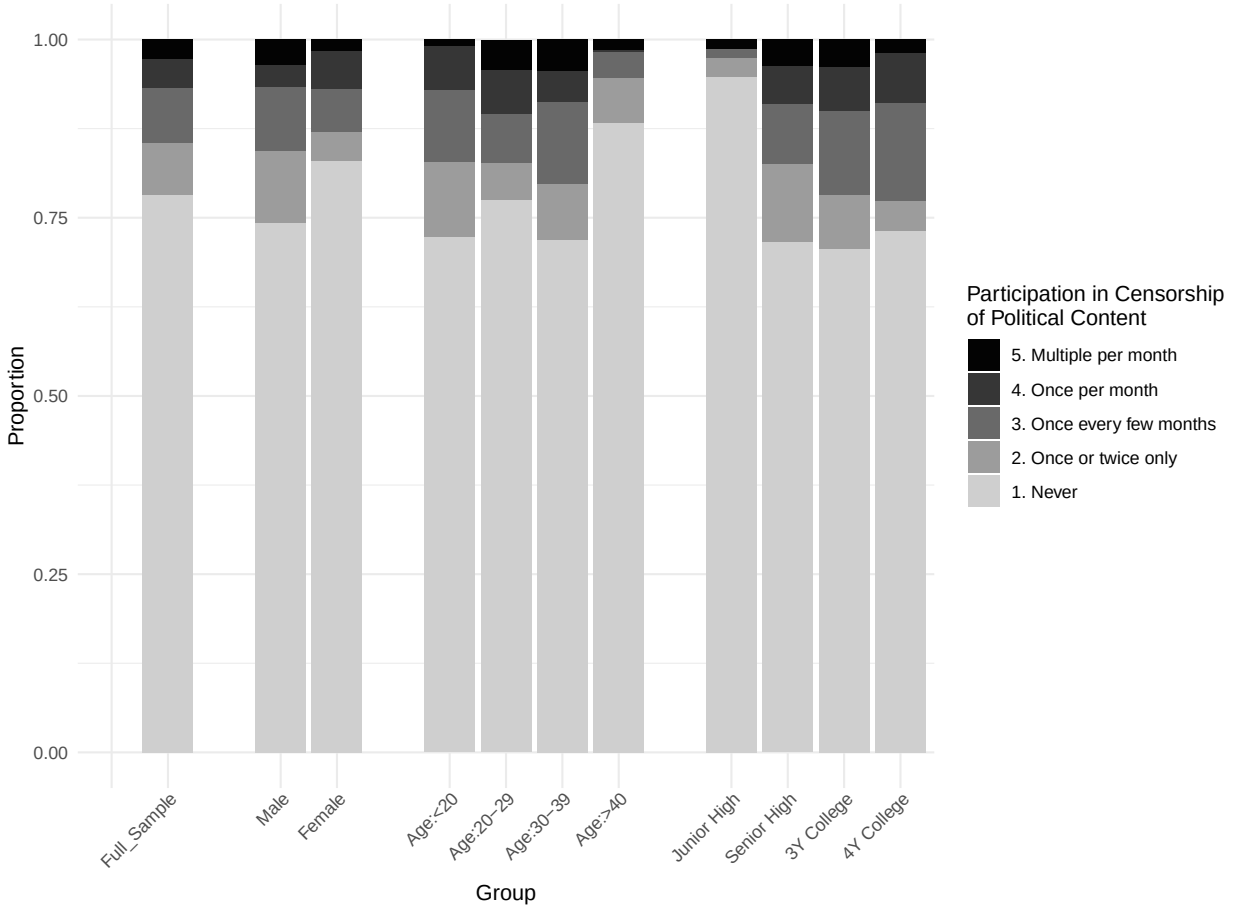
Note: Data about Chinese Internet users are from *The 45^{th} Statistical Report of Internet Development in China*, issued by China Internet Network Information Center in April 2020.

Appendix C Study 1: Prevalence of Participation

C.1 Participation in Censorship of Specific Content Categories

Figure C1 reports the distribution of self-reported participation in the censorship of political content. In general, around 50% of the "participating respondents," or 25% of all respondents, self-report having participated in the censorship of political content. Men, younger generations, and the better-educated are significantly more likely to participate in the censorship of political content.

Figure C1: Distribution of Self-Report Participation in Censorship of Political Content

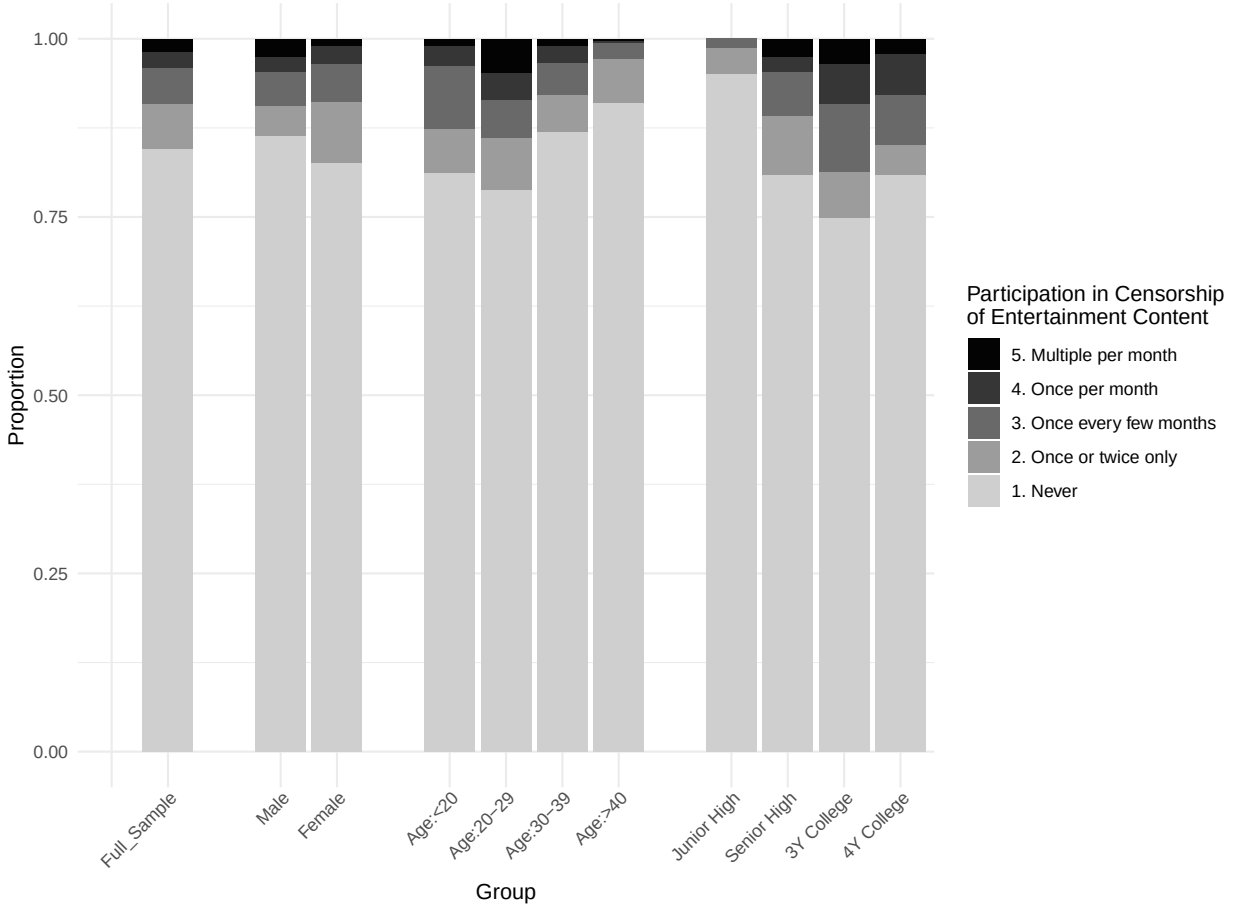


Note: All observations are weighted by gender, rural/urban location, region, age group, and education.

Figure C5 reports the distribution of self-reported participation in the censorship of entertainment content. Around one-third of the "participating respondents," or one-sixth of all respondents, self-reporting having participated in the censorship of entertainment content. In contrast to political content, females are more likely to report entertainment content than males.

Consistent with the political content, younger and better-educated are more likely to report entertainment content.

Figure C2: Distribution of Self-Report Participation in Censorship of Entertainment Content



Note: All observations are weighted by gender, rural/urban location, region, age group, and education.

C.2 Unweighted Sample

The unweighted sample shows a slightly higher proportion of respondents self-reporting participation in censorship.

Figure C3: Distribution of Self-Report Participation in Censorship: Unweighted Sample

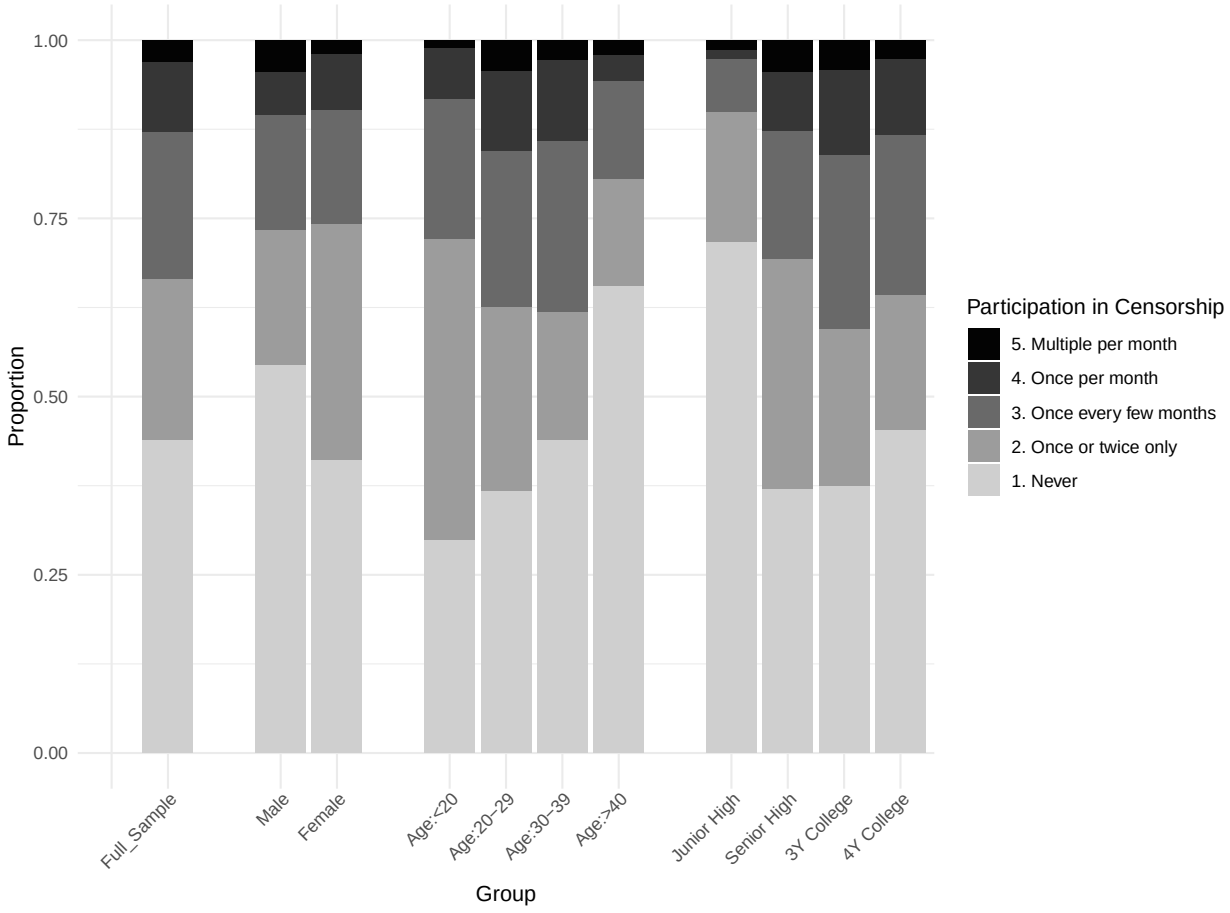


Figure C4: Distribution of Self-Report Participation in Censorship of Political Content: Unweighted Sample

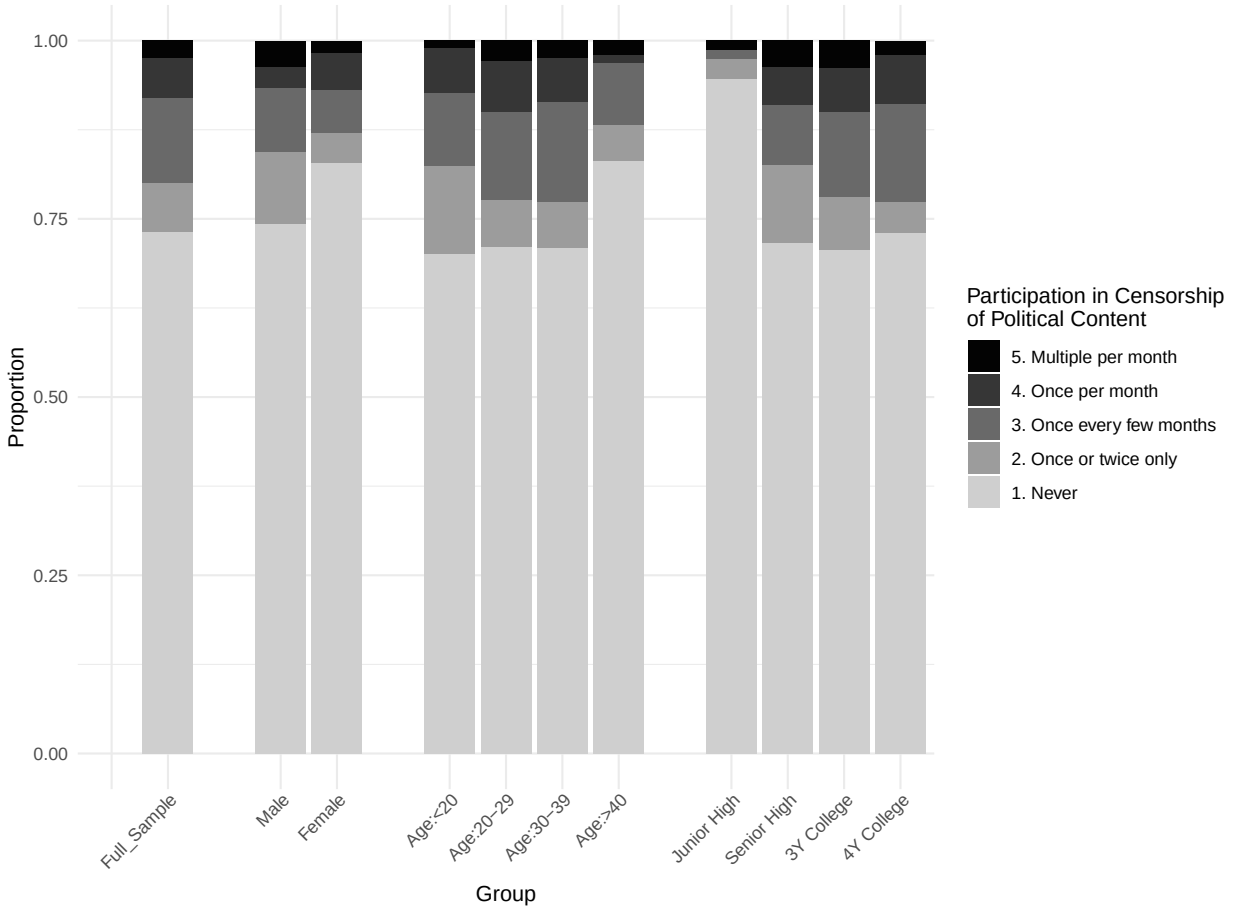
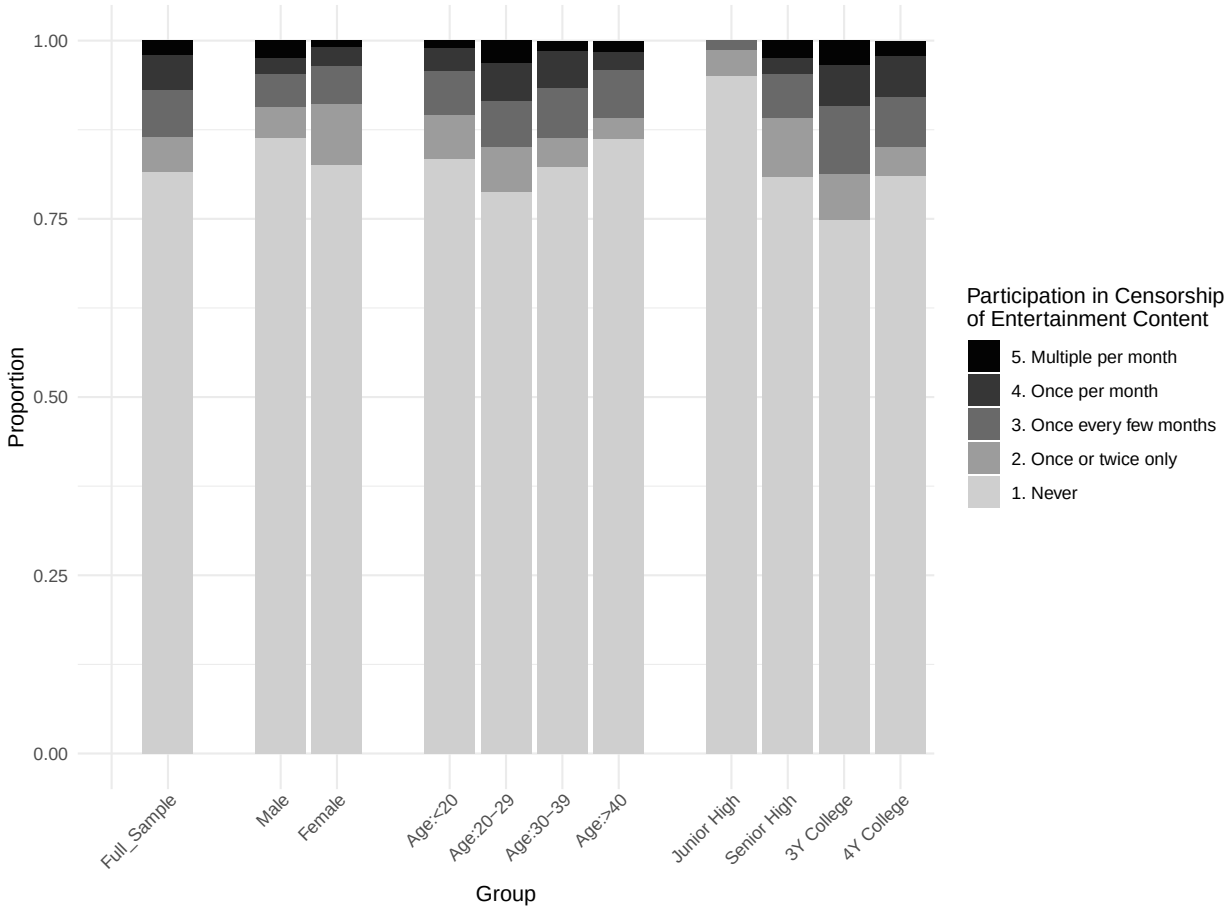


Figure C5: Distribution of Self-Report Participation in Censorship of Entertainment Content: Unweighted Sample



Appendix D Study 1: Correlation with Support

D.1 Main Analyses

Table D1: Correlation between Participation in Censorship and Support for Censorship Using the Five Point Measure of Participation

	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)
Female	0.179***	0.198***	−0.031	−0.046	0.012	0.0001
	(0.064)	(0.063)	(0.066)	(0.067)	(0.073)	(0.073)
Age Group	0.156***	0.127***	0.052*	0.031	0.104***	0.084***
	(0.027)	(0.027)	(0.028)	(0.028)	(0.031)	(0.031)
Education	−0.062**	−0.068**	0.060*	0.059*	0.036	0.035
	(0.030)	(0.031)	(0.031)	(0.032)	(0.034)	(0.035)
Urban	0.214***	0.225***	0.375***	0.338***	0.304***	0.286***
	(0.067)	(0.067)	(0.070)	(0.071)	(0.077)	(0.078)
Party Member		0.391***		−0.001		0.120
		(0.097)		(0.103)		(0.113)
Pol. Ideology		−0.003		−0.112***		−0.138***
		(0.025)		(0.026)		(0.029)
Econ. Ideology		0.202***		0.036		0.031
		(0.030)		(0.032)		(0.035)
Pol. Interest		−0.026		0.070***		0.093***
		(0.022)		(0.023)		(0.025)
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)
Weighted Sample	✓	✓	✓	✓	✓	✓
N	1,088	1,071	1,084	1,066	1,086	1,068
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 Likert scale. Standard errors in parentheses. Participation in censorship is measured on a five-point scale: never participated, once or twice only, once per few months, once per month, and multiple times per month.
*p < .1; **p < .05; ***p < .01

Table D2: Correlation between Specific Types of Participation in Censorship and Support for Censorship Using the Five Point Measure of Participation

	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)
Female	0.197***	0.187***	−0.036	−0.056	−0.003	−0.004
	(0.063)	(0.063)	(0.067)	(0.067)	(0.073)	(0.073)
Age Group	0.110***	0.108***	0.017	0.010	0.077**	0.076**
	(0.026)	(0.026)	(0.027)	(0.027)	(0.030)	(0.030)
Education	−0.063**	−0.061**	0.059*	0.067**	0.038	0.039
	(0.031)	(0.031)	(0.032)	(0.032)	(0.035)	(0.035)
Urban	0.223***	0.222***	0.332***	0.330***	0.286***	0.286***
	(0.068)	(0.068)	(0.071)	(0.071)	(0.078)	(0.078)
Party Member	0.402***	0.411***	−0.001	0.009	0.129	0.131
	(0.097)	(0.097)	(0.103)	(0.104)	(0.113)	(0.113)
Pol. Ideology	0.001	0.004	−0.107***	−0.104***	−0.137***	−0.137***
	(0.025)	(0.025)	(0.026)	(0.026)	(0.029)	(0.029)
Econ. Ideology	0.207***	0.207***	0.041	0.040	0.033	0.033
	(0.030)	(0.030)	(0.032)	(0.032)	(0.035)	(0.035)
Pol. Interest	−0.026	−0.024	0.070***	0.074***	0.093***	0.093***
	(0.022)	(0.022)	(0.023)	(0.023)	(0.025)	(0.025)
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)
Weighted Sample	✓	✓	✓	✓	✓	✓
N	1,071	1,071	1,066	1,066	1,068	1,068
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 Likert scale. Standard errors in parentheses. The independent variables are participation in political censorship and participation in censorship of entertainment and cultural content. Both independent variables are measured on a five-point scale: never participated, once or twice only, once per few months, once per month, and multiple times per month.

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

D.2 Robustness Checks

First, I transform the independent variable into a binary measure of participation. If the respondent has never participated before, I code it as 0, otherwise, I code it as 1. Table F10 shows the results using the binary measure of participation to re-run the analyses. As demonstrated in the table, the relationship between participation and support for censorship remains the same using the binary measure.

Table D3: Correlation between Participation in Censorship and Support for Censorship Using Binary Measure of Participation

	Support for Censorship		Support for Censorship of Political Content		Support for Censorship of Non-Political Content	
	(1)	(2)	(3)	(4)	(5)	(6)
Participate (Binary)	0.184***	0.189***	0.184***	0.214***	−0.135*	−0.103
	(0.064)	(0.063)	(0.066)	(0.067)	(0.073)	(0.073)
Female	0.165***	0.191***	−0.040	−0.053	0.008	−0.008
	(0.063)	(0.063)	(0.066)	(0.067)	(0.073)	(0.073)
Age Group	0.155***	0.133***	0.056**	0.035	0.081**	0.059*
	(0.028)	(0.028)	(0.028)	(0.029)	(0.032)	(0.032)
Education	−0.056*	−0.065**	0.064**	0.064**	0.045	0.044
	(0.030)	(0.030)	(0.031)	(0.032)	(0.034)	(0.035)
Urban	0.225***	0.240***	0.386***	0.355***	0.297***	0.280***
	(0.067)	(0.068)	(0.070)	(0.071)	(0.077)	(0.078)
Party Member		0.421***		0.030		0.133
		(0.097)		(0.103)		(0.112)
Pol. Ideology		−0.007		−0.117***		−0.132***
		(0.025)		(0.027)		(0.029)
Econ. Ideology		0.204***		0.038		0.035
		(0.030)		(0.032)		(0.035)
Pol. Interest		−0.031		0.064***		0.095***
		(0.022)		(0.023)		(0.025)
Constant	2.697***	2.123***	2.594***	2.669***	2.498***	2.504***
	(0.125)	(0.177)	(0.129)	(0.186)	(0.143)	(0.205)
Weighted Sample	✓	✓	✓	✓	✓	✓
N	1,088	1,071	1,084	1,066	1,086	1,068
Adjusted R^2	0.019	0.106	0.046	0.088	0.020	0.088

Notes: Dependent variables are indicated in column headings and are measured on a five-point Likert scale. Standard errors in parentheses. Participation is a binary variable indicating whether the respondent has participated before.
*p < .1; **p < .05; ***p < .01

Second, the observational analyses may be susceptible to the possibility of omitted variable bias or reverse causality. To mitigate these concerns, I conducted a sensitivity analysis following Cinelli and Hazlett (2020) to test the robustness of the main model to potential unobserved confounders or reverse causal relationships. The results of the analysis, shown in column 5 of Table D4 (RV), suggest that a potential unobserved confounder or a reverse causal arrow through such a confounder would need to explain at least 8.8% of the residual variance of both the treatment and the outcome to explain away the estimated treatment effect. In comparison, an unobserved confounder as strong as *Economic Ideology*, the most significant predictor in the main model, can only explain 4.2% of the residual variance. Therefore, the model is at least robust to an omitted variable as strong as the most significant covariate in the current model.

Table D4: Sensitivity of the Main Regression Model to Unobserved Confounders

Treatment:	Outcome: *Support for Censorship*					
	Estimate	Std. Error	t-value	$R^2_{Y\sim D\|\mathbf{X}}$	RV	$RV_{\alpha=0.05}$
Participation in Censorship	0.084	0.028	2.997	0.8%	8.8%	3.1%
df = 1061; *Bound (Z as strong as Economic Ideology)*: $R^2_{Y\sim Z\|\mathbf{X},D}$ = 4.2%, $R^2_{D\sim Z\|\mathbf{X}}$ = 0.4%						

Notes: RV stands for robustness value, the proportion of residual variance of both treatment and outcome a confounder needs to explain in order to explain away the treatment effect. $RV_{\alpha=0.05}$ is the RV such that the treatment is no longer statistically significant at 0.05 level. $R^2_{Y\sim D|\mathbf{X}}$ is the proportion of residual variance of treatment that a confounder needs to explain in the extreme scenario that it explains 100% of residual variance of the outcome Cinelli and Hazlett (2020). The benchmark is *Economic Ideology*, the most significant predictor in the main model.

D.3 Additional Mechanisms

An additional mechanism that I tested is the increase in the perceived benefit of censorship. Specifically, participation allows ordinary users to report content they disapprove of, thereby increasing their perceived benefit. When censorship is solely a top-down process imposed upon ordinary users, they are more likely to have cynical views of the censorship apparatus and perceive themselves as victims of censorship. Conversely, because of the increased perceived empowerment, individuals are more likely to view government censorship activities as enforcing their censorship preferences. From the perspective of ordinary users, their participation redefines the government's role as an arbitrator of public demand on the Internet, rather than a manipulator of public opinion. Hence, they are more likely to view censorship as a tool they can use to suppress political opponents, increasing their support for censorship.

Hypothesis: As individuals participate more in the censorship process, they are more likely to believe that censorship benefits ordinary citizens such as themselves, which subsequently leads to greater levels of support for government censorship.

To examine whether participation increases the perceived benefit of censorship, the survey asked respondents about whether ordinary people are the victims or the beneficiaries of the current censorship apparatus. Using OLS regression models with all relevant covariates and adjusted by sample weights, I find no evidence to support the perceived benefit mechanism ($\beta = -0.077$, $p = 0.191$). The insignificant results may be due to the lack of variation in responses, which are concentrated on the middle choices. This suggests that respondents may not have a strong opinion on the perceived benefit question.

Appendix E Study 2: Experimental Design & Randomization Check

E.1 Simulated Social Media Posts

Figure E1 shows an example of the simulated social media posts. In the control groups (Upper Panel), there are three buttons under each post: "Like," "Share," and "Comment." In both treatment groups (Lower Panel), there are four buttons, a "Report" button (the button on the right) in addition to the three in the control group.

Figure E1: Simulated Social Media Posts

E.2 Question Wording

Table E1: Measurement of Main Outcome Variables

Hypothesis	Survey Items	Expectation
Censorship Support	Do you agree or disagree: The government should actively control the Internet and remove content that it deems inappropriate.	+
	Do you agree or disagree: The government should actively control online discussions on **government policies and party leadership**, and remove content that it deems inappropriate.	+
	Do you agree or disagree: The government should actively control online discussions on **entertainment stars and popular culture**, and remove content that it deems inappropriate.	+
Regime Support	How satisfied are you with the overall situation in China right now?	+
	Both the central and the local governments of all levels always work for the people and serve their needs.	+
	I completely trust both the central and the local governments.	+

E.3 Balance Table & Randomization Check

Table E2: Balance Table (Group Mean & F-test)

	Control	Treatment 1	Treatment 2	p-value
Female	0.483	0.484	0.497	0.465
Age Group	2.757	2.765	2.748	0.830
Education	2.987	3.043	3.001	0.686
Party Member	0.162	0.166	0.144	0.216
Economic Ideology	3.791	3.796	3.806	0.662
Region	2.855	2.872	2.846	0.879
Nationalism	4.444	4.398	4.443	0.970
Political Interest	3.536	3.590	3.565	0.513
Social Media Usage	3.721	3.708	3.733	0.759
Foreign Connection	0.831	0.801	0.821	0.781

Table E3: Randomization Check: Using Covariates to Predict Treatment

	Group Assignment	
	Treatment 1 − Control	Treatment 2 − Treatment 1
Female	−0.003	0.017
	(0.020)	(0.020)
Age	−0.003	0.002
	(0.010)	(0.010)
Education	0.016	−0.014
	(0.012)	(0.012)
Party Member	0.006	−0.045
	(0.028)	(0.029)
Ideology	0.003	−0.002
	(0.012)	(0.012)
Region	0.002	−0.003
	(0.006)	(0.006)
Nationalism	−0.017	0.022*
	(0.013)	(0.013)
Political Interest	0.012	−0.006
	(0.009)	(0.009)
Social Media	−0.004	0.006
	(0.010)	(0.010)
Foreign	−0.022*	0.013
	(0.012)	(0.012)

Appendix F Study 2: Analyses

F.1 Overall Results

Comparing the Control Group and Treatment Group 1

Table F1: The Effect of Providing the Opportunity to Participate 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)
Treatment 1	0.067*	0.069**	0.066*	0.074**	0.028	0.029
	(0.036)	(0.035)	(0.037)	(0.036)	(0.038)	(0.038)
Female		−0.086**		0.007		−0.028
		(0.036)		(0.037)		(0.039)
Age		0.127***		0.100***		0.087***
		(0.018)		(0.019)		(0.019)
Education		0.003		0.017		−0.001
		(0.021)		(0.022)		(0.022)
Party Member		0.091*		0.048		0.069
		(0.049)		(0.051)		(0.053)
Ideology		0.272***		0.267***		0.245***
		(0.021)		(0.022)		(0.022)
Nationalism		0.116***		0.071***		0.140***
		(0.023)		(0.024)		(0.025)
Political Interest		0.036**		0.011		0.040**
		(0.016)		(0.017)		(0.017)
Social Media		−0.014		−0.030		−0.042**
		(0.018)		(0.019)		(0.019)
Foreign		−0.058***		−0.038*		−0.029
		(0.021)		(0.022)		(0.023)
Constant	3.491***	1.603***	3.468***	1.886***	3.679***	1.953***
	(0.026)	(0.151)	(0.026)	(0.158)	(0.027)	(0.163)
N	2,664	2,504	2,668	2,507	2,662	2,501
Adjusted R^2	0.001	0.128	0.001	0.095	−0.0002	0.092

Notes: Dependent variables are indicated in column headings and are measured on a five-point Likert scale. Standard errors in parentheses. Only the blank control and control groups are included in the analyses.
*p < .1; **p < .05; ***p < .01

Comparing Treatment Groups 1 and 2

Table F2: Intention-To-Treat Effect of the Encouragement Treatment

	Support for Censorship		Support for Censorship of Political Content		Support for Censorship of Non-Political Content	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment 2	0.086**	0.076**	0.068**	0.057*	0.062*	0.059
	(0.034)	(0.033)	(0.034)	(0.034)	(0.037)	(0.036)
Female		−0.031		0.045		−0.029
		(0.034)		(0.034)		(0.037)
Age		0.105***		0.078***		0.095***
		(0.017)		(0.017)		(0.019)
Education		−0.013		0.023		−0.003
		(0.020)		(0.020)		(0.022)
Party Member		0.075		0.058		0.024
		(0.048)		(0.049)		(0.053)
Ideology		0.220***		0.211***		0.213***
		(0.019)		(0.020)		(0.021)
Nationalism		0.117***		0.084***		0.138***
		(0.022)		(0.022)		(0.024)
Political Interest		0.057***		0.043***		0.067***
		(0.015)		(0.016)		(0.017)
Social Media		−0.012		−0.041**		−0.048***
		(0.017)		(0.017)		(0.019)
Foreign		−0.038*		−0.039*		−0.044**
		(0.020)		(0.020)		(0.022)
Constant	3.473***	1.755***	3.466***	1.993***	3.645***	1.866***
	(0.053)	(0.151)	(0.054)	(0.155)	(0.058)	(0.167)
N	2,647	2,493	2,653	2,499	2,645	2,492
Adjusted R^2	0.002	0.106	0.001	0.081	0.001	0.093

Notes: Dependent variables are indicated in column headings and are measured on a five-point Likert scale. Standard errors in parentheses. Only the treatment and control groups are included in the analyses.
*p < .1; **p < .05; ***p < .01

F.2 Instrumental Variable Analysis

Main Analysis

Table F3: 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)
Report Click	0.228***	0.219***	0.202***	0.199***	0.130**	0.123**
	(0.054)	(0.052)	(0.055)	(0.054)	(0.058)	(0.056)
Female		−0.050*		0.027		−0.022
		(0.028)		(0.029)		(0.031)
Age		0.146***		0.116***		0.110***
		(0.015)		(0.015)		(0.016)
Education		0.004		0.023		0.003
		(0.016)		(0.017)		(0.018)
Party Member		0.097**		0.070*		0.053
		(0.040)		(0.042)		(0.043)
Ideology		0.258***		0.248***		0.238***
		(0.016)		(0.017)		(0.018)
Nationalism		0.092***		0.064***		0.121***
		(0.019)		(0.019)		(0.020)
Political Interest		0.032**		0.016		0.050***
		(0.013)		(0.013)		(0.014)
Social Media		−0.005		−0.031**		−0.049***
		(0.014)		(0.015)		(0.015)
Foreign		−0.048***		−0.046***		−0.043**
		(0.017)		(0.017)		(0.018)
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)
N	3,990	3,764	3,997	3,770	3,989	3,763

Notes: 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

Robustness Check

To check the robustness of the treatment effect, I use an alternative measurement of participation in censorship: the number of times the respondents clicked a "Report" button. As shown in Table F4, consistent with the main analyses, additional clicking of the "Report" buttons induced by the treatments significantly increases support for censorship.

Table F4: 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)
Report Click #	0.080***	0.077***	0.071***	0.070***	0.046**	0.043**
	(0.019)	(0.018)	(0.020)	(0.019)	(0.020)	(0.020)
Female		−0.046		0.031		−0.019
		(0.029)		(0.030)		(0.031)
Age		0.151***		0.121***		0.113***
		(0.015)		(0.016)		(0.016)
Education		0.007		0.026		0.004
		(0.017)		(0.017)		(0.018)
Party Member		0.098**		0.071*		0.054
		(0.040)		(0.042)		(0.043)
Ideology		0.259***		0.249***		0.239***
		(0.016)		(0.017)		(0.018)
Nationalism		0.091***		0.064***		0.121***
		(0.019)		(0.020)		(0.020)
Political Interest		0.033**		0.017		0.050***
		(0.013)		(0.013)		(0.014)
Social Media		−0.009		−0.035**		−0.052***
		(0.014)		(0.015)		(0.015)
Foreign		−0.050***		−0.048***		−0.044**
		(0.017)		(0.018)		(0.018)
Constant	3.484***	1.641***	3.463***	1.901***	3.673***	1.926***
	(0.024)	(0.122)	(0.025)	(0.126)	(0.026)	(0.131)
N	3,990	3,764	3,997	3,770	3,989	3,763

Notes: Report click number is the number of the "Report" buttons that the respondents clicked on the simulated social media page. All individual survey items were measured on a five-point scale.
LATE = Local Average Treatment Effect
*p < .1; **p < .05; ***p < .01

Effects by Treatment

I further disaggregate the CACE for Treatments 1 and 2 (Tables F5 and F6). The results show that, while both treatments significantly 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.

Table F5: Complier Average Causal Effects (CACE) of Treatment 1 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)
Report Click	0.156*	0.161**	0.154*	0.171**	0.065	0.068
	(0.085)	(0.081)	(0.087)	(0.085)	(0.089)	(0.088)
Female		−0.087**		0.006		−0.028
		(0.036)		(0.037)		(0.039)
Age		0.133***		0.108***		0.090***
		(0.018)		(0.019)		(0.020)
Education		0.006		0.020		0.001
		(0.021)		(0.022)		(0.022)
Party Member		0.105**		0.062		0.075
		(0.050)		(0.052)		(0.054)
Ideology		0.271***		0.266***		0.244***
		(0.021)		(0.022)		(0.022)
Nationalism		0.110***		0.065***		0.137***
		(0.023)		(0.024)		(0.025)
Political Interest		0.035**		0.010		0.040**
		(0.016)		(0.017)		(0.017)
Social Media		−0.016		−0.032*		−0.043**
		(0.018)		(0.019)		(0.019)
Foreign		−0.060***		−0.041*		−0.030
		(0.021)		(0.022)		(0.023)
Constant	3.491***	1.617***	3.468***	1.901***	3.679***	1.959***
	(0.026)	(0.150)	(0.026)	(0.157)	(0.027)	(0.162)
N	2,647	2,493	2,653	2,499	2,645	2,492

Notes: 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

Table F6: Complier Average Causal Effects (CACE) of Treatment 2 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)
Report Click	0.417**	0.382**	0.330*	0.289*	0.299*	0.295
	(0.169)	(0.170)	(0.169)	(0.172)	(0.180)	(0.185)
Female		−0.031		0.045		−0.029
		(0.034)		(0.035)		(0.037)
Age		0.152***		0.113***		0.131***
		(0.027)		(0.027)		(0.029)
Education		−0.002		0.032		0.006
		(0.021)		(0.021)		(0.023)
Party Member		0.120**		0.092*		0.059
		(0.053)		(0.054)		(0.058)
Ideology		0.220***		0.211***		0.212***
		(0.020)		(0.020)		(0.021)
Nationalism		0.080***		0.056*		0.109***
		(0.028)		(0.029)		(0.031)
Political Interest		0.051***		0.038**		0.063***
		(0.016)		(0.016)		(0.017)
Social Media		−0.018		−0.046***		−0.053***
		(0.018)		(0.018)		(0.019)
Foreign		−0.053**		−0.051**		−0.056**
		(0.022)		(0.022)		(0.024)
Constant	3.379***	1.728***	3.392***	1.971***	3.579***	1.845***
	(0.092)	(0.159)	(0.092)	(0.160)	(0.098)	(0.173)
N	2,647	2,493	2,653	2,499	2,645	2,492

Notes: 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

Effects by Type of Posts

Additionally, I analyze respondents' reporting behavior based on whether they clicked "report" under a pro-government post or an anti-government post. 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.

Table F7: Complier Average Causal Effects (CACE) of Reporting Anti-Regime Posts 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)
Report Anti	0.243***	0.233***	0.215***	0.212***	0.138**	0.131**
	(0.058)	(0.056)	(0.059)	(0.058)	(0.061)	(0.060)
Female		−0.051*		0.026		−0.023
		(0.028)		(0.029)		(0.031)
Age		0.146***		0.116***		0.110***
		(0.015)		(0.015)		(0.016)
Education		0.005		0.024		0.003
		(0.017)		(0.017)		(0.018)
Party Member		0.098**		0.071*		0.054
		(0.040)		(0.042)		(0.043)
Ideology		0.259***		0.249***		0.239***
		(0.016)		(0.017)		(0.018)
Nationalism		0.091***		0.064***		0.121***
		(0.019)		(0.020)		(0.020)
Political Interest		0.032**		0.016		0.049***
		(0.013)		(0.013)		(0.014)
Social Media		−0.007		−0.032**		−0.050***
		(0.014)		(0.015)		(0.015)
Foreign		−0.047***		−0.046***		−0.043**
		(0.017)		(0.017)		(0.018)
Constant	3.484***	1.655***	3.463***	1.913***	3.673***	1.934***
	(0.024)	(0.121)	(0.025)	(0.125)	(0.026)	(0.130)
N	2,647	2,493	2,653	2,499	2,645	2,492

Notes: 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

Table F8: Complier Average Causal Effects (CACE) of Reporting Pro-Regime Posts on Support for Censorship

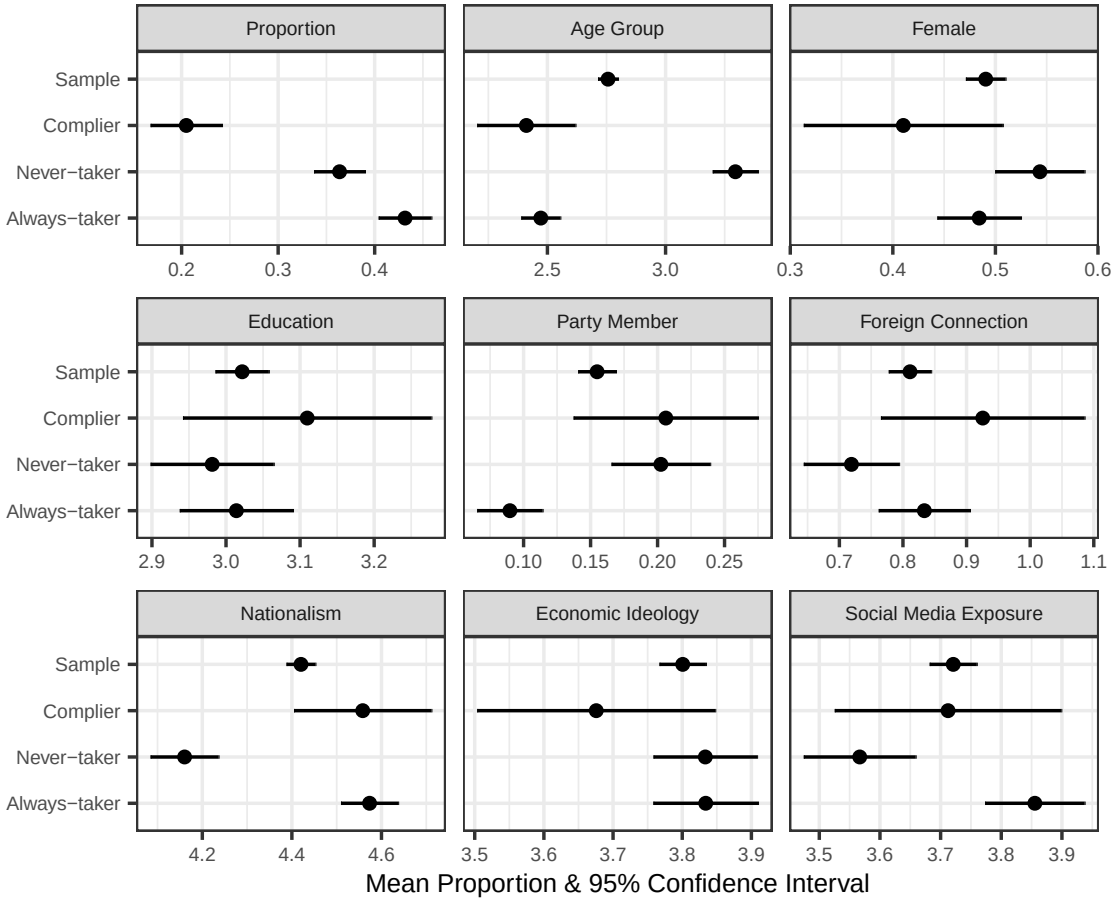
	Support for Censorship		Support for Censorship of Political Content		Support for Censorship of Non-Political Content	
	(1)	(2)	(3)	(4)	(5)	(6)
Report Pro	0.642***	0.615***	0.563***	0.551***	0.371**	0.353**
	(0.152)	(0.147)	(0.154)	(0.150)	(0.159)	(0.155)
Female		−0.046		0.032		−0.019
		(0.029)		(0.030)		(0.031)
Age		0.139***		0.110***		0.107***
		(0.015)		(0.015)		(0.016)
Education		−0.0001		0.020		0.0005
		(0.017)		(0.017)		(0.018)
Party Member		0.090**		0.064		0.050
		(0.041)		(0.042)		(0.043)
Ideology		0.254***		0.244***		0.236***
		(0.017)		(0.017)		(0.018)
Nationalism		0.102***		0.073***		0.127***
		(0.019)		(0.019)		(0.020)
Political Interest		0.033**		0.017		0.050***
		(0.013)		(0.013)		(0.014)
Social Media		0.001		−0.025*		−0.045***
		(0.014)		(0.015)		(0.015)
Foreign		−0.051***		−0.049***		−0.044**
		(0.017)		(0.018)		(0.018)
Constant	3.486***	1.632***	3.465***	1.892***	3.673***	1.920***
	(0.024)	(0.124)	(0.024)	(0.127)	(0.025)	(0.131)
N	2,647	2,493	2,653	2,499	2,645	2,492

Notes: 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

F.3 Profiling Compliers

Following Marbach and Hangartner (2020), I plot out the characteristics of compliers, always-takers, and never-takers, assuming there are no defiers. Figure F1 shows individuals who clicked the "Report" buttons tended to be younger and more familiar with social media. In contrast, the non-compliers, those who never click the "Report" buttons, tended to be older, nationalists with few foreign connections and limited social media exposure.

Figure F1: Profiling Compliers, Never-Takers, and Always-Takers Using Treatment Groups 1 & 2



Note: The first panel indicates the estimated proportion of compliers and never-takers. The remaining eight panels demonstrate the estimated group means of the full sample, compliers, and never-takers across eight different pre-treatment covariates. CI = Confidence Interval

F.4 Regime Support

Consistent with the findings in Study 1, I do not find significant effects of participation treatment on regime supports. This provides additional support for the cognitive dissonance theory and the system justification theory as explained in the main text in section 3.4.

Comparing the Control Group and Treatment Group 1

Table F9: The Effect of Providing the Opportunity to Participate on Regime Support

	Regime Satisfaction		Regime Assessment		Regime Trust	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment Group 1	0.011	0.032	0.002	0.015	0.031	0.060*
	(0.032)	(0.031)	(0.032)	(0.031)	(0.033)	(0.032)
Constant	4.066***	2.220***	4.046***	2.185***	4.023***	1.996***
	(0.023)	(0.135)	(0.022)	(0.133)	(0.023)	(0.137)
Covariates		✓		✓		✓
N	2,636	2,477	2,636	2,476	2,645	2,487
Adjusted R^2	−0.0003	0.111	−0.0004	0.118	−0.00005	0.113

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

Comparing Treatment Groups 1 & 2

Table F10: Intention-To-Treat Effect of Encouragement Treatment on Regime Support

	Regime Satisfaction		Regime Assessment		Regime Trust	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment Group 2	0.056*	0.034	0.051	0.033	0.029	0.005
	(0.032)	(0.032)	(0.031)	(0.031)	(0.032)	(0.031)
Constant	4.020***	2.502***	3.998***	2.378***	4.024***	2.357***
	(0.051)	(0.145)	(0.050)	(0.141)	(0.050)	(0.143)
Covariates		✓		✓		✓
N	2,616	2,465	2,618	2,466	2,628	2,479
Adjusted R^2	0.001	0.082	0.001	0.098	−0.0001	0.086

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

References

Cinelli, Carlos and Chad Hazlett. 2020. "Making Sense of Sensitivity: Extending Omitted Variable Bias." Journal of the Royal Statistical Society Series B-Statistical Methodology 82(1):39–67.