TIKTOK, SPIRAL OF SILENCE, AND ADMINISTRATIVE EXPANSION: INSIGHTS FROM A FIELD EXPERIMENT IN CHINA

Yiying Li[*] Meng Miao[†]

this draft: Dec 2024.

Abstract

How do short video platforms like TikTok influence public opinion? In this study, we explore how exposure to nationalist short videos—content that promotes national pride, supports government policies, and denounces opposition—shapes participants' willingness to express dissent. Conducted during the reform of a Chinese law aimed at enhancing government administrative powers, our survey experiment reveals that just one minute of exposure to these immensely popular videos significantly reduces the likelihood of submitting opposing comments, even when anonymity is guaranteed. This effect is primarily driven by the 'spiral of silence,' where viewers overestimate public support for nationalism and fear being isolated or labeled as traitors by the majority if they voice dissenting opinions. The effect is particularly pronounced in regions with a history of political movements and among wealthier individuals. We estimate that short videos contribute to a nearly 60% reduction in opposing comments, highlighting TikTok's powerful role in shaping public sentiment and enabling the expansion of administrative powers.

Keywords: Social media, Short video, Institutional change, The spiral of silence, Nationalism
JEL Classification: D72, L82, P16

[*]School of Finance, Renmin University of China. Emails: Yying.li@ruc.edu.cn, miaomeng@ruc.edu.cn, respectively.

[†]We are very grateful to David Stromberg for his valuable suggestions on the model and theoretical sections.

1 Introduction

The explosive growth of short video platforms like TikTok has significantly reshaped how individuals consume information and share opinions. By 2024, adults in the U.S. are expected to spend 58.4 minutes a day on TikTok, over twice the 27.4 minutes they spent in 2019.[1] As people increasingly turn to short videos for quick content delivery—even on serious issues like politics—concerns have grown over their role in fragmenting information, spreading misinformation, and polarizing political views (Wang, 2020; Zhu *et al.*, 2020).[2] In numerous countries, particularly authoritarian ones, these media outlets are used to manipulate public opinion, possibly creating more significant threats to democracy, the freedom of speech and the idea of free market. This paper investigates how short video platforms, due to their high efficiency in shaping public opinion, may silence opposition during legislative changes aimed at increasing executive power, possibly leading to institutional setbacks.

We hypothesize that short video platforms influence public opinion through the "spiral of silence" effect (Noelle-Neumann, 1974). This effect describes situations that individuals exposed to media tend to overestimate the extent of support for a prevailing view, leading them to hesitate in expressing their dissenting opinions out of fear of isolation.[3] The "spiral of silence" is based on two key assumptions: first, that people frequently misinterpret public opinion, perceiving media depictions as the dominant viewpoint; and second, that individuals tend to withhold opposing views if they believe those views con-

[1] Based on the statistical data and forecast report published by Statista on February 13, 2024.

[2] The addictive nature of short videos stems from their ability to stimulate the brain's default mode network (DMN) and trigger dopamine release, reinforcing pleasurable experiences through simple swiping (Su *et al.*, 2021). Algorithmic advancements further exacerbate this by tailoring content to individual preferences, deepening user engagement and fostering addiction.

[3] Recent economic studies have increasingly emphasized the critical role that agents' perceptions of others play in shaping economic and political outcomes. Like Bursztyn *et al.* (2020a, 2023a,b), Furnas & LaPira (2024), and see Bursztyn & Yang (2022) for a review of more related works. Meanwhile, many studies have raised concerns about the negative consequences of media control and censorship in China, particularly regarding political cognitive biases, such as Qin *et al.* (2018), Chen & Yang (2019) and Luo & Li (2024).

flict with mainstream sentiment.[4] Short videos, with their authentic grassroots hosts and repetitive themes, may be more powerful than traditional media in manipulating public opinion and leading the spiral of silence effect by amplifying biases in how people perceive public consensus (Reisach, 2021; Gao *et al.*, 2023).

To examine the potential for a spiral of silence effect, we conducted a randomized controlled trial (RCT) in China—the birthplace of TikTok and the world's largest short video market. Our study focuses on a significant event: the proposed amendments to China's Public Security Administration Punishment Law in September 2023. These amendments sought to empower the government to penalize vaguely defined actions that "hurt national sentiment," sparking widespread concern over the potential expansion of executive power. During the public consultation period from September 1 to 30, 2023, approximately 100,000 individuals submitted over 120,000 comments to China's National People's Congress, making it one of the most controversial and widely debated drafts. Meanwhile, TikTok and other platforms were flooded with videos stressing the defense of national sentiment, often employing remarkably similar rhetoric. Alarmingly, those in favor of revising or removing the law were labeled "traitors," "western obsession," or "kowtowing to foreigners."[5]

To evaluate how nationalistic videos influence public attitudes, we conducted an online experiment during the law's public consultation period through an online survey. Over 2,000 participants from across China were randomly assigned to different treatment and control groups. Each participant in the treatment groups watched one or two representative 1-minute videos sourced from TikTok. These videos fell into two categories: the first, a "nationalism clip," supported the new legal provisions from a nationalism perspective, emphasizing the necessity of protecting national sentiment; the second, a "empowerment clip," encouraged participants to take a stand against the amendments by

[4] The spiral of silence suggests that the silence of dissenters becomes a self-fulfilling prophecy, as their silence further skews the perceived balance of opinions, leading to more silence. However, this process is influenced by various factors. Our focus is on a key stage of this process—how people interpret others' views based on the information they receive, and how this reduces the expression of dissenting opinions.

[5] While there is no direct evidence connecting these videos to specific interest groups, the uniformity of language suggests possible coordination. This, however, is beyond the scope of this paper.

highlighting concerns over vague language and the potential for unchecked administrative power. After watching the videos, participants were asked to submit their opinions—either in support of or against the amendments—with the assurance that their responses would be anonymously submitted to the official consultation website.

The findings reveal that even a one-minute video can significantly shift attitudes and behavior. Participants exposed to the empowerment clip were 7.19% more likely to submit revision suggestions than the control group (which watched no videos), while those who watched nationalist videos were 5.28% less likely to submit. Furthermore, viewing incendiary nationalist videos that insulted critics of the law—labeling them as "traitors"[6]—further reduced the likelihood of submitting revision suggestions by 7.55%, underscoring how negative framing and fear of social ostracism can suppress dissent.

In addition to the "spiral of silence" effect, short videos may also influence behavior by directly persuading individuals, aligning their beliefs with the narrative presented in the video. We employed two methods to separate this persuasion effect from the spiral of silence. First, we asked participants to estimate how *others* would react to opposing comments, thereby measuring their perceived social pressure. We find that those who view nationalist videos significantly overestimate this pressure, more likely believing that others would view opposing comments as "foreign sympathizing", or "traitor". This perceived pressure leads to behavior changes: participants who overestimate social intolerance are notably less likely to submit opposing suggestions. These findings confirm that short videos can create a sense of social pressure, effectively suppressing the expression of opposing views.

To further distinguish the impact of social pressure from belief revision and confirm a causal link between participants' attitude shifts and the spiral of silence, we carried out a supplementary experiment with 536 participants.[7] Participants in this group were shown nationalist videos and subsequently provided with accurate information about the true

[6] Currently, China does not have specific laws governing the news media format of short videos, which allows for the complete absence of constraints on such defamation of others.

[7] In the baseline experiment, participants were more tolerant of opposition than they assumed others to be. On average, they rated their own belief that opposition is 'worshipping foreign cultures' at 5.14, but estimated others' views at 5.69, indicating an overestimation of social intolerance.

level of social intolerance among others, as established by the baseline experiment. Relative to the baseline group, who viewed nationalist videos without being informed of others' actual attitudes, these respondents exhibited an 8.30 percentage point higher probability of submitting opposing views for each unit of overestimation on a 1-to-10 scale.[8] This finding indicated that the nationalist videos led participants to overestimate social intolerance by 0.39 points, which in turn reduced opposing opinions by 3.24% (=0.39*8.30%). This reduction accounted for 61.36% (=3.24/5.28%) of the overall reduction in opposing opinions caused by the nationalist videos—an economically significant effect, given that the baseline probability of submitting a revision suggestion in the control group was only 11.70%.

Reluctance to voice opposition may not only come from the spiral of silence but also from other factors. In our pilot study, 113 participants identified three primary reasons for not submitting feedback: 1) the view that the expansion of government power would not affect them, as long as they complied with the law, 2) the ability to free-ride on the efforts of others, and 3) the belief that their opinions would not be taken into account by authorities. In response to these concerns, we provide all participants in the baseline experiment with the necessary information to address these aspects, ensuring a consistent informational background for all participants when making feedback decisions. We also consider the potential impact of fears related to identity exposure and government retaliation, as well as the tendency for individuals to project self-intolerance onto social intolerance. Our empirical analysis indicates that these factors have minimal influence on our results.

Through our analysis of heterogeneous effects, we find that historical movements that weakened social trust are instrumental in shaping the spiral of silence, primarily by causing individuals to overestimate social intolerance toward dissent. Our analysis examines the institutional framework shaped by significant historical events in China, including the Cultural Revolution (1966-1976), the One-Child Policy, and the Great Famine. During the Cultural Revolution, citizens were encouraged to report on each

[8] We informed participants of the actual average attitudes of others toward opposition, based on a 1-to-10 scale where 1 represents absolute tolerance and 10 represents absolute intolerance.

other, eroding family bonds and undermining social trust. A similar pattern is observed in areas with strict enforcement of the One-Child Policy, where the norm of reporting violations became entrenched, as well as in regions affected by the Great Famine, where a deep sense of survival-driven social distrust emerged. Our findings show that participants from these regions perceive higher levels of social intolerance after watching nationalist videos, suggesting that historical experiences of distrust intensify the spiral of silence, leading to greater self-censorship.

We further analyze how individual traits influence the second stage of the spiral of silence—how much participants chose to silence themselves in response to perceived social pressure. Our findings indicate that wealthier and more educated individuals are more inclined to self-censor under the same perceived social pressure.[9] While these participants don't overestimate social intolerance, they are more inclined to alter their behavior under similar pressure, likely as a strategy to protect their wealth and status from potential criticism. This significantly affects public discourse: well-informed individuals who generally support limiting government power often withdraw, thereby diminishing the checks on authority when they are most crucial to curbing the expansion of administrative power.

To estimate the real-life aggregate impact of video viewing on the reduction of opposing comments during public consultations, we apply our experimental results while considering two key factors: 1) real-world challenges to commenting, such as issues with website accessibility, and 2) the reduced effectiveness of videos for viewers who watch them repeatedly. On short video platforms, nationalism-themed videos outnumber empowerment videos by nearly five to one. Our calculations indicate that short videos led to a reduction of 188,985 opposing comments, which is a 60% drop. This dramatic decline in opposition suggests a lack of significant pushback from the public, potentially influencing the vague wording in the second draft of the bill, which made actions like "wearing or displaying symbols that glorify invasion wars" punishable.

What is the long-term effect of short videos, particularly nationalist ones, on shaping

[9] A similar trend is observed in participants from regions severely affected by the Cultural Revolution.

participants' self- and perceived social intolerance, as well as on their choice to submit feedback? Our follow-up experiments conducted 6 and 12 months later reveal that these effects are negligible. The appearance of new nationalist videos rapidly decreased to zero after the public comment period ended on September 30, leading to a sharp decline in public interest in the legal amendment. In the follow-up surveys, we find that most respondents could not recall the content of the new provision (90%) or the videos they had watched (82%). Unsurprisingly, only a small fraction still expressed ongoing concerns about the legal amendment. Moreover, the effects of the videos, especially the nationalist ones, had completely dissipated, with all groups showing the same 16% likelihood of submitting feedback. This indicates that using short videos as a tool to influence public attitudes carries little risk of backlash, as the long-term effects of such media are virtually negligible.

To the best of our knowledge, our paper is the first to quantify the impact of the spiral of silence on specific policy formulation and reform. The impact of the spiral of silence is widely recognized in media research, with the majority of studies showing that public opinions are shaped by media coverage as a result of the fear of social isolation (Gunther, 1998; Scheufele *et al.*, 2001; Gearhart & Zhang, 2018). Factors like the specific media issues discussed, the context in which the information is received, the emotional tone of media coverage, and cultural influences such as Confucian traditions, all play a significant role in shaping the spiral of silence (Huang, 2005; Gerber & Rogers, 2009; Askay, 2015; Gearhart & Zhang, 2018; Matthes *et al.*, 2018). The emergence of participatory social media has been seen as a potential disruptor of the spiral of silence effect (Laor, 2024). However, our study, along with the findings of Huang & Ho (2023), indicates that the spiral of silence remains a prevalent phenomenon in political discourse. More specifically, our study has confirmed the internalization of social norms, demonstrating that the spiral of silence can be guided by symbols such as nationalism, even in the context of increasingly diverse content and anonymous opinion expression on emerging media platforms. Additionally, Clemente & Roulet (2015) has explained the impact of spiral of silence in the realm of institutional development based on qualitative method. In this study, We employ RCT

method and quantify psychological and behavioral reactions to confirm the existence of the spiral of silence in institutional development.

Our study aligns with the literature on the public's misperception of others. People often adjust their actions to conform to societal expectations or avoid disapproval from others (Miller *et al.*, 1999; Gelfand *et al.*, 2024). However, People's perceptions of others often exhibit biases, and these misperceptions can influence their behavior in various aspects, such as financial investment, consumption habits, and voting (Allcott & Rogers, 2014; Perez-Truglia & Cruces, 2017; Kast *et al.*, 2018; Kalwij *et al.*, 2019). In recent years, research on misperceptions has increasingly focused on the domain of social norms. For example, Cantoni *et al.* (2019) demonstrates that individuals' behavior in collective action contexts is influenced by their concerns about others' opinions. When individuals are made aware of others' real beliefs, their actions change significantly, highlighting the role of perceived social norms in shaping behavior. Similarly, in the study of Bursztyn *et al.* (2020b), Saudi Arabian husbands tend to overestimate society's disapproval of their wives working outside the home, even though they have no objection to their own or others' wives pursuing external employment. This misperception is widespread and contributes to the social norm discouraging wives from working outside the home. Our study explores how, in the context of institutional reform and national sentiments, short videos contribute to the misperception of others. These misperceptions often lead to the suppression of dissent against unjust policies, as individuals adjust their behavior to align with perceived societal norms.

2 Institutional Background

2.1 Public Opinion Feedback In the Chinese Legislative Process

China initiated public participation in its legislative process in 2005 by posting the draft Property Law online for public input. The feedback period typically extends for one month, allowing citizens to submit anonymous feedback through the National People's

Congress (NPC) website.[10] Individuals can offer multiple comments, often each addressing specific clauses of the draft. While the total comment count is visible, individual comments remain private unless commenters opt to publish their feedback on other platforms.[11] Once the feedback period concludes, only the draft and the number of submitted comments are displayed.

In April 2008, the NPC Standing Committee's Meeting institutionalized public participation by mandating that all draft laws be published online to collect public input. Since then, it has become standard procedure to invite public opinion on proposed legislation. To date, the NPC website records over 160 public consultations, many of which involved multiple revisions. Laws that closely impact daily life, such as the Labor Contract Law, the Law on the Protection of Women's and Children's Rights, and the Civil Code, have drawn significant public engagement. For example, the 2018 revisions to the Civil Code attracted 440,000 comments from 110,000 participants.

After public feedback is collected, the legislative process requires an explanation of which clauses were amended based on public consultation, especially those with substantial comments. It also mandates justifications for any clauses that were commented on but left unchanged. In practice, however, the feedback details are rarely disclosed with the final legislation, leaving contributors unsure whether their suggestions were seriously considered.

2.2 The Revision of the "Public Security Administration Punishment Law"

Within China's legal framework, the "Public Security Administration Punishments Law" is of considerable significance, covering a wide range of behaviors similar to those addressed by the "Criminal Law", but for less severe offenses. Unlike the "Criminal Law", this law follows a simplified procedure, allowing public security agencies to impose

[10] Chinese citizens can submit their opinions on proposed amendments to major laws, such as the "Public Security Administration Punishments Law" in this study, via the NPC website (www.npc.gov.cn). Occasionally, certain laws are handled by other agencies, such as the Ministry of Civil Affairs, which gathers feedback on "Marriage Law" amendments.

[11] When amendments are controversial, some citizens and lawyers may choose to post their submitted comments on social media sites like Weibo and TikTok to signal their stance and rally similar feedback from the public.

penalties, including detention of up to one month, without requiring involvement from the procuratorate or judiciary. If a penalty is later deemed erroneous, individuals may seek state compensation only after paying their fine and serving their sentence. The law provides for vaguely defined offenses like "picking quarrels and provoking trouble," enabling authorities to use broad charges to detain individuals for activities such as petitioning, with detentions of up to one month.

Seventeen years after its initial enactment, the "Public Security Administration Punishments Law" was revised for the first time in 2023, with public feedback collected on the National People's Congress website from September 1 to September 30. Key areas in the draft include amendments on information security and privacy rights, but the most widely discussed change is the addition of Article 34, which aims to prevent actions that disrespect national heroes or harm national sentiments. Specifically, it outlines penalties for: 1) forcing or persuading others to wear or display attire or symbols in public that damage the spirit of the Chinese nation or offend national sentiments, and 2) creating, distributing, or promoting items or statements that insult or harm national sentiments. Violations may lead to detention of five to ten days and fines of up to 5,000 RMB (or 700 USD).

This amendment sparked widespread debate online, as Chinese social media saw a surge of articles and videos expressing completely contradictory opinions. Legal professionals caution that phrases like "damage the spirit of the Chinese nation" and "hurt national sentiments" are too abstract and vague, leaving them open to subjective interpretation by law enforcement. This ambiguity, they warn, could result in excessive discretion and even abuses of power. At the same time, pro-nationalist videos endorsing these terms are widely circulated, often pushing extreme nationalist messages while overlooking the need for precise and scientific legal standards. Supporters of these videos label critics as overly anti-government, accusing them of sowing distrust in law enforcement. A striking observation is that videos advocating for , both in terms of the number of videos and the number of views, far surpass those promoting empowerment. This imbalance may exert a substantial influence on the shifting public sentiment and the eventual outcomes of the

legal amendments.

3 Study Framework

3.1 Theoretical Framework

In this section we present a simple model to illustrate how biased videos can lead to misperceptions about others, influencing public behavior. Defining y_i as the tendency of individual i to remain silent and refrain from commenting:

$$y_i = y_i^0 + \beta \overline{y} \tag{1}$$

where y_i^0 represents an individual's initial intention to stay silent, and $\overline{y}$ is the average attitude across all individuals. The parameter β represents the extent to which individuals are influenced by social pressure. A higher β indicates stronger social pressure and a greater likelihood of considering others' views. The motivation for considering others' opinions can stem from several factors. One reason is self-censorship, as individuals may fear punishment if their behavior y_i differs too much from the social consensus $\overline{y}$. Another is the "wisdom of the crowds": if many people believe something is right, it is more likely correct.

Through continuous social learning and the updating of information, an individual's belief about the expectations of others' behavior should align with the true probability of those behaviors, that is:

$$\overline{y} = \tfrac{1}{n} \textstyle\sum_{i=1}^{N} y_i \tag{2}$$

which leads to the equilibrium solution:

$$\overline{y} = \tfrac{1}{1-\beta} \tfrac{1}{n} \textstyle\sum_{i=1}^{N} y_i^0, \tag{3}$$

where $\frac{1}{n} \sum_{i=1}^{N} y_i^0$ represents the average initial tendency to remain silent among individuals. The multiplier, $\frac{1}{1-\beta}$ captures the leverage effect of social pressure β. As social pressure β increases, the multiplier grows, indicating that individuals give more weight to

others' opinions, making them more likely to remain silent in equilibrium. In the extreme case, when β approaches 1, the tendency to stay silent becomes nearly infinite, regardless of one's initial inclination to speak. This reflects the fear of expressing opposing views or the societal pressure that forces everyone to remain silent.

In China, media outlets frequently display significant bias, often influenced by propaganda, and may deliberately lead audiences to adopt beliefs in fabricated or overstated public opinions (Qin *et al.*, 2018). Moreover, the algorithms that recommend videos based on user preferences further amplify the bias in the news individuals are exposed to. The video shock may have two types of effects on the belief of individual i. First, the persuasion effect occurs when the individual becomes more aligned with the video's content, leading to a change in their own motivation: $y_i^0 \to \theta y_i^0$, where θ is greater than 1 for nationalism videos that encourage silence, and less than 1 for empowerment videos that foster dissent.

The second impact is that the video shock could result in a misperception of the average public: $\overline{y} \to \alpha\overline{y}$, where α is greater than 1 for nationalism videos, which cause individuals to overestimate the social pressure against writing opposing comments, and less than 1 for empowerment videos, which foster an estimate of social pressure that encourages such comments.

The likelihood for proposing comments becomes:

$$y_i^{video} = \theta y_i^0 + \beta\alpha\overline{y} \tag{4}$$

And the equilibrium solution becomes:

$$\overline{y}^{video} = \frac{\theta}{1-\beta\alpha}\frac{1}{n}\sum_{i=1}^{N} y_i^0, \tag{5}$$

Compared with original $\overline{y}$, the change in the coefficient ($\frac{1}{1-\beta} \to \frac{\theta}{1-\beta\alpha}$) illustrates that both the persuasive effect and the spiral of silence effect are present. The persuasive effect is evident in that when everyone watches the nationalist video, each individual's opinion increases by a factor of θ, which in turn raises the average level of silence by

the same factor. For the spiral of silence effect, the original β parameter is replaced by $\beta\alpha$, indicating that the increase in individuals' misperceptions of others' silence, captured by α, has a similar impact to social pressure, β, in fostering silence. This framework reflects the classic theory of the spiral of silence, with α and β representing two stages: α refers to the initial stage, the misperception of others' attitudes induced by short videos, while β describes the fear individuals experience when expressing dissenting views due to perceived social pressure.

3.2 Experimental Design

From September 1 to September 30, 2023, the NPC solicited public feedback on the draft amendments to the Public Security Administration Punishment Law through its official website. During this period, we ran an intervention experiment on China's largest professional online survey platform. The intervention consisted of three parts: a pilot, a baseline, and a supplementary experiment, involving 3,030 participants and resulting in 2,861 valid questionnaires. The detailed experimental procedure is outlined in Figure 1.

In the baseline experiment, respondents are randomly assigned to one of four main video intervention groups:

Control Group: Participants in this group do not watch any related videos.

Nationalism Group: Participants in this group watched videos that support the law from a nationalist viewpoint, discouraging opposing proposals. These videos are drawn from the top three TikTok clips in the topic related to this legal revision, based on the number of likes for each type.[12] To further assess how video content shapes attitudes, we selected two videos: The *incendiary* clip shows an emotional speaker denouncing opponents as traitors, while the *calm* version presents a measured speaker refuting the claim that "hurting national sentiments" lacks a clear definition. Participants were randomly assigned to watch one of those videos.

Empowerment Group: Participants in this group watched videos that focus on

[12] We curated representative nationalist videos on TikTok, characterized by over 90% script homogeneity and high engagement (over 300,000 "likes" on top-viewed videos). Videos were re-edited to mask the publisher's identity to eliminate additional biases.

the legal aspects of the provision, critiquing the its disregard for legal principles and its arbitrary approach. The videos suggested that the provisions are intentionally vague to permit broad discretionary power, raising concerns about government overreach and potential abuse of authority.[13]

Mixed Group: Participants in this group watched both a nationalist video and a empowerment video, with the order of presentation randomized.

This setup allowed us to analyze the differential impacts of nationalist and empowerment video content, as well as the combined effect of both, on respondents' attitudes and behaviors toward the proposed legal provisions. Table A1 outlines the themes and representative phrases from each video used in the intervention.

Differences in respondents' knowledge base or perceptions of essential information may significantly impact their inclination to submit an opposing proposal. For example, a respondent who doubts their input will impact legislative decisions may feel less motivated to participate. To identify these belief-shaping factors, we conducted a pilot test involving 113 participants (see Figure 1). After watching the videos, participants explained their hesitations about suggesting amendments. The three primary reasons identified were as follows: 1) They felt that as long as they complied with the law, the expansion of administrative power would not negatively affect them, 2) They could benefit from the efforts of others without contributing themselves, and 3) They believed their opinions would go unheard. In response, we implemented three text-based information interventions in the baseline experiment that : 1) highlighting the dangers of power abuse, 2) emphasizing the public's general enthusiasm for proposing amendments, and 3) providing examples of successful public-driven legal changes. Full content details are available in Appendix A.

To track shifts in respondents' attitudes, we asked them to estimate their *likelihood* of making comments at various stages of the questionnaire: before and after watching the

[13] Similar to our approach with nationalist videos, empowerment videos are divided into two types: one type uses strong rhetoric to criticize the arbitrariness of those in power, while the other takes a weaker stance, emphasizing how vague legal provisions could leave enforcement guidelines unclear. Our experimental results suggest that the intense empowerment videos produce a more notable intervention effect.

videos, and after each information shock. At the end of the baseline experiment, following all three information interventions, we invited respondents to write down any specific suggestions for the provision. To encourage honest feedback, we informed respondents that their answers would remain anonymous and that we would submit their suggestions to the official website on their behalf.[14] If they had no suggestions or comments, they could leave the response blank. Finally, we collected basic demographic information, including respondents' hometown, age, gender, education, income, occupation, and attitudes toward money, trust, and cooperation.

To ensure the reliability of the data, we excluded responses from the video-watching group that were completed in under 3 minutes, as recorded by the survey platform. We also removed responses where participants consistently selected the same option across all questions (e.g., always choosing the first option). Our baseline experiment began with a sample size of 2,206, resulting in 2,122 valid responses after exclusions.

In a supplementary experiment, we engaged 536 additional respondents with the nationalist videos and introduced new information interventions to help disentangle the effects of personal beliefs and the spiral of silence channel. While holding other conditions constant, we provided participants with information on how the public perceives submitting amendments as a traitorous act, as reflected in the reaction scores from prior respondents in the baseline experiment. The behavioral changes from this information were then used to measure the extent of the spiral of silence effect. We conducted random follow-ups with respondents from the baseline experiment six months and one year later.

4 Analysis of Experimental Results

4.1 Sample Description

We randomly assigned respondents to different groups to minimize the impact of potential confounding factors. Table 1 displays descriptive statistics for the full sample and each major subgroup's performance, along with T-tests comparing each treatment

[14] We cannot allow direct submission on the legislative website, as we are unable to access the content provided by participants, given that all comment submissions are kept confidential.

group to the control group. The table shows that the main treatment groups (those who watched only the nationalist or empowerment video) do not differ significantly from the control group in gender, age, education, income, pre-intervention nationalism, and three life attitude measures.

We also map the provincial distribution of respondents, as shown in Figure A1 in Appendix A. The map indicates that respondents are randomly distributed across almost all provinces in mainland China.

4.2 Baseline Results

4.2.1 Group Performance

This study's main outcome variable is whether respondents ultimately share their opinions and the nature of those opinions. We categorize respondents' views into two types: first, "Opposing comments", where respondents oppose the provision, citing concerns about administrative overreach or ambiguous wording, and suggest deletion or amendment. Second, "Supporting comments", where respondents fully support the provision, with some recommending stronger penalties, such as "increasing detention days." No respondents expressed both support and opposition at the same time.

Figure 2 shows the percentage of respondents in each group who wrote comments expressing either opposition (Dark gray bars) or support (light gray bars). Opposing comments are least frequent among respondents in the nationalist video group (6.42%) and most frequent in the empowerment video group (18.89%). The mixed video group (12.96%) and the control group (11.70%) fall between these two extremes. Respondents exposed to the incendiary nationalist video are even less likely to write opposing comments than those who viewed the calm nationalist video, underscoring the stronger suppressive effect of a harshly critical tone that brands dissenters as traitors. Unsurprisingly, the pattern for supportive comments goes in the opposite direction, indicating that the video has a contrasting effect on whether people express support or opposition.

4.2.2 Baseline Empirical Model

Our regression analysis examines average marginal effects using a Logit model. The dependent variable $1_{opposing,i}$ is a binary indicator, which takes the value 1 if respondent i wrote an opposing comment, and 0 otherwise. The main independent variables $VideoType_i$ are binary indicators representing different video interventions: V_{Nat_i} (whether the respondent watched the nationalist video or not) and V_{Emp_i} (whether the respondent watched the empowerment video or not). We control for respondents' basic characteristics, θ_c, including age, income, gender, education level, and whether they were aware of the amendment before participating in the survey. In addition, we include province-fixed effects ϵ_p.

$$Logit\left(P\left(1_{opposing_i}=1\right)\right)=\beta_0+\beta_1\times V_{\text{type}_i}+\theta_c+\epsilon_p+\mu_i, \tag{6}$$

As shown in Table 2, viewing nationalist videos significantly reduce respondents' likelihood of writing opposing opinions, whereas empowerment videos have the opposite effect. In Column (1), the average marginal effect of V_{Nat} is -0.0643, significant at the 1% level, implying that exposure to only nationalist content decreases the probability of writing an opposing comment by 6.43 percentage points. Given that the average likelihood of writing an opposing opinion is 12.49 percentage points, this translates into a notable (6.43/12.49=) 48.52% reduction—an economically meaningful impact. The effect for respondents who are exposed to incendiary nationalist content are even more pronounced. In Column (2), the average marginal effect for $V_{Nat:Inc}$ is -0.0873, suggesting that exposure to incendiary nationalist content reduces the likelihood of writing an opposing opinion by 8.73 percentage points, a result that is statistically and economically significant.

We examine whether the videos influence the tone of commenters' responses by applying two different sentiment analysis methods. The first method adopts a *bag-of-words* approach, where comments containing words with negative or emotional connotations (e.g., "abuse," "treason"), according to a Chinese lexicon, are classified as aggressive

and coded as 1, while comments without such language are coded as 0.[15]The second method employs *SnowNLP*, a natural language processing tool, to conduct semantic analysis, assigning each opinion a sentiment score, where higher scores indicate a more negative tone after dealing (Li *et al.*, 2022).[16]

Columns (3) and (5) in Table 2 provide the regression outcomes on the emotional tone of opposing comments using bag-of-words and SnowNLP method, respectively. Among respondents who voice opposition, viewers of the nationalist video typically maintain a moderate tone, whereas those who watched the empowerment video express more intense emotion and dissatisfaction, indicating strong discontent with the new law. This suggests that the video treatments influence both the likelihood of writing opposing comments and the intensity of those responses. Again, the effect is more pronounced for incendiary nationalism videos; See Columns (4) and (6) in Table 2.[17]

Additionally, we conduct a regression analysis to explore how nationalist videos affect the likelihood of writing supporting (instead of opposing) opinions and the emotional tone of these comments. The results, presented in Table B1 in Appendix B, demonstrate that nationalist videos notably raise both the likelihood of supportive responses and the intensity of emotional tone, further validating the significant impact of nationalist content on individuals' behavior and emotions.

[15] We constructed a dictionary for analyzing the opinions expressed by respondents based on the NTUSD Chinese Sentiment Dictionary published by National Taiwan University, and then combined it with Jieba for word segmentation and a Naive Bayes bag-of-words model to evaluate each review, where a score of 1 indicates negative sentiment while 0 for neutral and -1 for positive. Details can be seen in Table A2.

[16] The sentiment analysis begins with text segmentation into individual words or phrases, followed by part-of-speech tagging to identify the roles of nouns, verbs, adjectives, etc. Next, a pre-trained sentiment model analyzes the sentiment, categorizing it as positive, negative, or neutral. Based on the prevalence and weight of sentiment words, an overall score is computed, with higher values indicating a more negative tone.

[17] In Tables C1 and C2, we demonstrate that our baseline results remain robust even after excluding comments with ambiguous attitudes and using another measurement of social intolerance.

5 Disentangling the Effects of the Spiral of Silence and Persuasion

In this section, we examine the mechanisms through which video interventions influenced respondents' behavior. Following the theoretical framework (Section 3.1), we hypothesize that the Spiral of Silence, along with individual belief updates (i.e. the persuasion effect), is instrumental in creating behavioral changes triggered by short video interventions. To test this, we incorporated questions into the survey to probe participants' views and their expectations of others' attitudes toward opposition—namely, whether expressing dissent against this law intended to protect "national sentiment" could be labeled as "traitorous."

- *Do you think proposing amendments to this legal provision is an act of being a "traitor"?*
- *Do you think that others might see proposing amendments to this legal provision an act of being a "traitor"?*

Both questions are scored on a scale from 1 to 10, with higher scores indicating a greater level of intolerance towards the act of proposing amendments.

We begin by testing the effects of video interventions on respondents' levels of self-intolerance and anticipated social intolerance. Table 3 shows that the nationalist videos significantly increase both self- and social intolerance. On average, exposure to these videos raises self-intolerance by 0.36 points and social intolerance by 0.39 points. Incendiary nationalist videos have the strongest impact, increasing self-intolerance by 0.80 points and social intolerance by 0.86 points, while calm nationalist videos have significant impact only on social intolerance (+0.25 points). In contrast, empowerment videos reduce both self- and social intolerance by 0.43 and 0.47 points, respectively. These findings suggest that video interventions have a profound effect on both personal beliefs and perceived societal views regarding opposition to the law.

We then explore how respondents' self- and social intolerance influence their decision to write opposition to the law. Figure 3 reveals a clear negative correlation between the probability of writing opposing opinions and both respondents' self- and social- intolerance. Table 4 presents a regression analysis examining the correlation between respondents' beliefs and the likelihood of writing opposing opinions. The negative average marginal effects suggest that as respondents' belief in viewing opposition as "traitor" sentiment strengthens, and as they perceive others share this belief, their likelihood of submitting opposing opinions decreases.[18]

However, these results are considered "Naive" because both self-beliefs and expectations of others are likely shaped by the same "shock" —exposure to nationalist video interventions. The shared shock may cause correlated errors in both beliefs, which undermines the reliability of the average marginal effect coefficients in Table 4 . This makes it difficult to distinguish between the spiral of silence effect and the possibility that respondents were convinced by the video, limiting any meaningful interpretation of their economic significance.

To disentangle the effects of the "spiral of silence" mechanism from "persuasion," we conducted a supplementary experiment focused solely on the nationalist video interventions. This setup replicated the baseline experiment with one significant change: after watching the nationalist video, participants were shown the extent of social intolerance drawn from our baseline experiment—an average score of 5.14 out of 10 from 2,212 respondents on whether proposing amendments might be viewed as a "traitor." This information aimed to enhance participants' understanding of social intolerance while minimally affecting personal beliefs. By comparing comment decisions between respondents with and without this additional information, we can better isolate the impact of social conformity pressure and perceived social intolerance on individual choices. Figure A2 illustrates the characteristics of respondents exposed to nationalist video interventions in our main and supplementary experiments. Both rounds show perfect random assignment,

[18] Table 4 shows that both perceptions have a positive effect on respondents' likelihood of expressing supportive opinions and leaving strengthening comments on the law, though the impact of social intolerance is not significant.

with the sample distribution remaining consistent across experiments.

Respondents provided estimates of self- and social intolerance twice—before and after the information intervention.[19] Figure 4 illustrates shifts in participants' perceptions of social intolerance after they received actual statistics on social intolerance. Respondents' estimates of others' intolerance shift significantly after the information intervention, with beliefs aligning closely to the provided statistics. Specifically, those who initially overestimate others' intolerance (above 5.14) decrease their estimates considerably, while those who underestimate (below 5.14) adjust upward upon learning the true values. In contrast, changes in their own intolerance levels toward opposing views remain minor, consistent with our assumption that self-intolerance and perceived social intolerance operate as distinct psychological accounts and are not necessarily linked.[20]

Regression analyses are conducted on a merged sample, which includes the subsample from our baseline experiment with the nationalism video intervention, along with the supplementary experiment sample. We construct a variable, *PerceptionGap*, to measure the gap between respondents' expectations of others' perception that opposition reflects "traitor" and the actual average perception (5.14). This variable captures whether respondents overestimate or underestimate social intolerance. Additionally, $Informed$ is defined as a dummy variable, taking the value of 1 if respondents are informed of respondents' actual average perception (5.14), and 0 otherwise.

According to the regression results of average effect based on OLS model in Column (1) of Table 5, holding other factors constant, those informed of the actual average perception have a 8.30% greater probability of writing opposing opinions for each point by which they previous overestimate social intolerance, on a 1-to-10 scale. This result further confirms that the spiral of silence effect: people are less likely to raise opposing

[19] On the other hand, comments on law modification were collected only after the intervention. We refrained from asking participants to comment on the legal provision in both instances before the intervention, as the additional effort required often leads to repetitive responses, thereby compromising data accuracy.

[20] It is worth noting that even after presented with accurate data, participants with initially higher (or lower) pre-intervention scores still tend to maintain higher (or lower) post-information assessments, indicating the persistence of their prior beliefs. See Figure A3.

comments under perceived social pressure from others' intolerance.[21]

Moreover, this result allow us to infer the proportion of the total effect of nationalism video intervention attributed to the spiral of silence mechanism.[22] From the empirical results in Table 3, we see that watching nationalist videos raises participants' expectations of others' intolerance by 0.39 relative to the control group. This implies that if participants' perceptions about others' attitudes are corrected to the level of the control group, the probability of revising their opinions would increase by 3.24% (=0.39×8.30%). Given that the probability of the nationalist intervention group expressing opposing opinions is 5.28% lower than that of the control group, approximately 61.36% (=3.24/5.28) of this reduction can be attributed to the pressure exerted by others—i.e., the spiral of silence. The remaining 38.64% of the observed behavior change is likely due to shifts in respondents' personal beliefs induced by the video intervention.

6 Excluding Alternative Interpretation

To ensure our interpretation of the results is not confounded by other factors, we examine several potential influences on participants' decision-making. We first introduce three additional text-based interventions for all respondents. In a pilot experiment prior to the main survey, we ask 113 respondents for reasons behind their reluctance to voice opposing opinions. Aside from personally supporting the new law, the main reasons include: 1) China's large population means others will comment, reducing their sense of responsibility. 2) They feel the legislative body is unlikely to adopt their input. 3) The legislation does not directly affect them because they do not wear unusual clothing.

In the main survey, we address each of these reasons through targeted informational interventions. To address reason 1, we inform participants of the number of opinions others

[21] Another piece of evidence from the empirical results is that receiving accurate information about others does not significantly increase the probability of writing supportive opinions. In the context of the nationalist video intervention, writing supportive opinions does not directly stem from external pressure (as analyzed in Table B2).

[22] Before being informed, the average expected intolerance of others is 6.27, dropping to 6.07 after information is provided, a decrease of 0.20. Meanwhile, respondents' self-intolerance only decreased by 0.003 after being informed. We can consider this impact as specifically targeting their perceptions of others, and thus the change in behavior is also due to the update in their views about others. In other words, the mechanism related to the spiral of silence can be disentangled through this impact.

have submitted to reduce the perception that their input is unnecessary. For reason 2, we share successful examples from the Women's Rights Protection Law, illustrating that the legislative body does consider public feedback. For reason 3, we provide cases where innocent individuals were implicated or even lost their lives due to abuses of power, highlighting the potential indirect effects of the legislation on everyone, as depicted in Table A1.

To prevent repetitive responses, we refrain from asking for written comments on the law after each intervention. Instead, we collect respondents' *willingness* scores (on a 1-to-10 scale) which reflect their likelihood to express opposing views following each informational intervention.[23] We find that informing respondents about the negative impacts of power abuse (reason 2) and successful precedents (reason 3) significantly increases their average willingness to propose opposing opinions. The information for reason 1 reduces respondents' willingness, likely because learning that thousands have already submitted opinions makes them feel their input is less impactful.

Despite the effects of the information interventions on respondents' willingness scores, Figure 5 shows that the ranking of willingness to express opposing opinions remains consistent across groups throughout the interventions, with nationalism showing the lowest and empowerment the highest willingness. This suggests that the initial video intervention's influence on participants' relative willingness to submit opposing comments outweighs the impact of factors such as a sense of responsibility or expectations about the legislative body's response.[24]

A common concern is that respondents' behavioral shifts may be influenced by fear of government authority or potential retaliation. The Chinese government's sustained push for nationalism, along with its tight media controls, may intensify this fear (Chen & Yang, 2019; Lu *et al.*, 2024). This concern is further heightened by the government's ability to trace individuals online through telephone numbers registered under real identi-

[23] There is a significant positive correlation between self-reported willingness and the actual submission of opposing opinions; See Table A3.

[24] Also the positive correlations between willingness scores at different stages are detailed in Table A3.

ties following real-name registration.[25] Respondents viewing the nationalism videos may interpret them as government-endorsed messages, prompting behavioral adjustments due to fear of retaliation.

We made several adjustments within the experiment to mitigate this concern. When embedding short-video content into the questionnaire, we removed the publishers' account information, preventing respondents from viewing it as government-endorsed. Additionally, we assessed respondents' perceptions of the government's stance on this legislative action by asking:

- *To what extent do you think the government will be unhappy if the law cannot be passed or requires major revisions?*

Higher scores reflect the belief that the government is actively promoting this legislation. We re-conduct the primary analysis, as presented in Table B3. After controlling for this government attitude rating, our main findings remain largely unchanged, suggesting that the behavioral shifts are not mainly driven by fear of government authority.

When analyzing the true impact of video interventions, we must consider another potential confounding mechanism: individuals' change in their exception social intolerance after viewing nationalism videos could be merely the projection of the shift of their own attitude. Previous research has demonstrated that individuals often project their own beliefs onto their perceptions of others' opinions (Wojcieszak & Rojas, 2011). This cognitive bias is evident across various contexts. For instance, when predicting changes in financial markets, respondents may assume that others share their expectations (Cohn *et al.*, 2015). Similarly, on video and media platforms, individuals tend to believe that their personal views are more widely held than they actually are. This projection can influence how they interpret information and engage with content, potentially reinforcing existing biases and shaping public discourse (Reisach, 2021; Gao *et al.*, 2023).

We provide several evidence that respondents' self intolerance and expected social intolerance could move independently. Firstly, the correlation between the two moder-

[25] The Real-Name Registration Reform, implemented in recent years, requires citizens to use their real identities online, enabling the government to quickly identify individuals making "inappropriate" remarks.

ate, with a coefficient of approximately 0.44 (panel A of Table B4). Moreover, in the supplementary experiment when individuals were informed with the true value of social intolerance, we find a much smaller increase in self intolerance, as suggested by Figure 4. Furthermore, when controlling for respondents' own tolerance levels, the empirical results remain consistent with our baseline regression, suggesting that this potential mechanism is unlikely to be the primary driver of our findings (results are displayed in the panel B of Table B4).

7 Heterogeneity Analysis

In this section, we explore the personal traits that make individuals more responsive to video interventions, especially in contexts where the spiral of silence effect is influential. We seek to clarify how these traits influence two steps of the spiral of silence: the extent to which information interventions shape respondents' views of social intolerance, and how those perceptions, in turn, affect their willingness to participate in commentary.

7.1 The Role in Shaping the Perception of Social Intolerance

The spiral of silence effect is primarily shaped by individuals' expectations of others' tolerance levels, which can be strongly influenced by collective memories of historical events. A notable example is the *Cultural Revolution*, during which people were encouraged to report, accuse, and even confront family members over alleged "counter-revolutionary ideology" (Walder & Su, 2003; Yan, 2024). This period resulted in widespread family breakdown and significant loss of life from 1966 to 1976, leaving a legacy of mistrust that remains today (Bai & Wu, 2023; Wang, 2021). Our hypothesis is that the first stage of the spiral of silence effect—namely, individuals' overestimation of intolerance following a nationalism video intervention—will be especially strong in regions that experienced higher levels of Cultural Revolution conflict. We measure this conflict intensity by the proportion of deaths during the Cultural Revolution relative to each municipality's 1964

population, capturing how social tolerance varied across regions.[26]

As shown in columns (1) and (2) of Table 6, respondents from regions with more Cultural Revolution victims perceive society as less tolerant after viewing nationalist videos. This highlights how collective memories of political trauma, like the Cultural Revolution, can disrupt social trust, foster suspicion of intolerance, and exacerbate the spiral of silence.

Two other significant historical events impacting social trust in China are the implementation of the *One-Child Policy* since 1982 and the experience of the *Great Famine* between 1959 and 1962. Public resistance to the One-Child Policy led local governments to encourage citizens to report on and expose one another, which ultimately weakened social trust (Otis, 2011; Shi & Kennedy, 2016; Cai & Feng, 2021). To capture the intensity of policy enforcement, we use provincial increases in one-child rates as reported in the 2000 Chinese Population Census (China's fifth national census).[27] Similarly, famine experiences can significantly shape individuals' views on social tolerance, with distrust arising from the scarcity of resources, competition for food, and government mismanagement during crises. (Meng *et al.*, 2015; Gooch, 2017).[28] The results in Table 6 suggest that both strict One-Child Policy enforcement and famine experiences are associated with a greater likelihood of respondents overestimating social intolerance in response to nationalist video interventions.

[26] This data, drawn from Chen *et al.* (2020), compiles historical records from approximately 3,000 local gazetteers across over 1,000 Chinese counties.

[27] We use the proportion of only children in marriageable families after the implementation of the family planning policy, minus the overall proportion of only children in the region, as a measure of the strictness of policy enforcement. The reason for using the difference is to eliminate potential interference from other factors. For example, in economically developed areas like Shanghai, the rate of only children has always been very high and consistent, not necessarily due to the influence of the family planning policy.

[28] According to the "Comprehensive Study of Major Natural Disasters and Mitigation Strategies in China" compiled by the National Science and Technology Commission's National Major Natural Disaster Research Group, the ten provinces most severely affected by disasters from 1958 to 1962 are Hebei, Liaoning, Jiangsu, Zhejiang, Anhui, Shandong, Henan, Hubei, Guangdong, and Sichuan. We assign a value of 1 to respondents from these provinces and 0 to others.

7.2 Role in Conforming to Social Intolerance

In this section, we focus on the impact of different factors on the relationship between the degree to which respondents overestimate others' intolerance (as measured by *PerceptionGap*, defined in Section 5) and the expression of opposing opinions. The goal of this process is to identify which respondents, characterized by certain traits, are more prone to conforming and reducing the expression of opposition when exposed to the same external pressures.

The column (3) of Table 6 shows that, on top of Cultural Revolution, both income and education strongly influence individuals' likelihood of expressing opposition. Wealthier and more educated individuals tend to stay silent, possibly because they fear that speaking out could put their wealth, reputation, or social standing at risk. In environments dominated by leftist ideology, affluent individuals are often targeted more intensely (Engesser *et al.*, 2017), which may encourage them to maintain a low profile. Furthermore, those with higher education may recognize from historical and personal experiences the political consequences that come with opposition, prompting adjustments in behavior. As those with greater social standing and resources tend to remain silent, it becomes harder to resist or stop legislation that strengthens government authority. We also conduct an analysis to understand how other respondent characteristics affect each step the spiral of silence; See Appendix B.3 for a detailed account.

8 Short- and Long-term Impacts of Video Interventions

8.1 Short-term Impacts

In reality, what impact do these online videos have on people's behavior in writing opposing comments, and how does this, in turn, affect the final outcomes of legislative amendments?

We searched for suitable short videos on China's two most popular short video

platforms, Tiktok and Kuaishou.[29] We selected videos that had titles or slogans about legal amendments, and had more than 1,000 likes.[30] Panel A of Table 7 provides a summary of the data of nearly 40 selected videos. There was a surge in the number of videos and a sharp increase in total views, likes, comments, and shares during the public comment period, which declined rapidly after September 30, when comments closed. Nationalism videos, based on total number of views, were far more popular than empowerment videos, likely to generate a "spiral of silence" where fear of social pressure fosters a one-sided narrative.

We use the outcomes of our experiments, particularly the baseline empirical estimate (the average marginal effects of video interventions $V_{\mathbf{Nat}}$ and $V_{\mathbf{Emp}}$ in the Table 2) in the likelihood of submitting amendment proposals, to assess the impact of these short videos on the total number of comments submitted to the public consultation. However, the respondents' actual likelihood of providing feedback may be driven by influences outside the scope of our experimental setting. First, by requesting feedback solely through our questionnaire, we made the process significantly easier than if participants had to submit feedback independently. In a real-life scenario, participants might not be aware of the option to comment on the legislation. Even if they were, they could face difficulty locating the feedback website or navigating it. Moreover, they may have heightened concerns about privacy breaches. Consequently, our results may overstate participants' true willingness to provide feedback.

Moreover, individuals are likely to come across multiple similar videos during the feedback window, as platforms' algorithms tend to recommend content that is similar to what users have already seen. It is reasonable to assume that the marginal effectiveness of video on commenting decreases as the number of views increases, meaning that the

[29] Based on the financial reports from the third quarter of 2024, TikTok's daily active users have exceeded 600 million, while Kuaishou's daily active user count stands at over 400 million.

[30] This represents a relatively high number of likes, given that anyone can upload videos on TikTok, yet few videos actually garner thousands of likes. Typically, the like-to-view ratio on TikTok ranges from 1% to 3%. Therefore, achieving 1,000 likes means the video likely has more than thirty thousands of views, which is not an easy feat. Details can be seen in the Figure D1 in Appendix D.

impact of a second viewing is less likely to be as strong as the first.[31]

We used various adjustment factors to account for their influence in real world scenario. For example, we assume that due to costs and information barriers, the likelihood of people voicing opposition is only one-fifth of the probability observed in our experiment. Additionally, repeated viewings would decrease the marginal willingness to submit feedback to 80% of the original level. Based on these assumptions, we conducted our estimations with corresponding robustness checks. Our findings, presented in Table 7, show that the imbalance in video viewings—where nationalist videos were five times more prevalent than empowerment videos—contributed to an average decrease of approximately 188,985 comments.[32] Therefore, even assuming that all 125,962 comments submitted to the legislative website were opposing opinions, our calculation suggests that the total number of opposing comments would have been reduced by 60% (=188,985 / (188,985 + 125,962)), a substantial reduction that must have a considerable impact on the legislative results.

The lack of sufficient critical feedback may be one reason why ambiguous clauses in these bills have not undergone substantial revision. In the latest version of the law, released in July 2024, certain previously vague statements, such as "hurting the feelings of the Chinese nation," were modified to read: "In public places... wear, display, or promote symbols or attire that glorify invasion wars or acts, causing adverse social impact." However, the revised clause still lacks precise definitions of prohibited actions, including terms like "promote" and "glorify." Furthermore, the second draft amendment in 2024 attracted minimal public attention, making it likely that the law will soon pass without significant revisions.

[31] Because TikTok and Kuaishou restrict view count data access to content creators, we estimated the view counts by analyzing the like numbers and applying like-to-view ratios, We have chosen a 5% like-to-view ratio for our evaluation, as this represents a very high engagement rate. This approach enables us to conservatively estimate the number of people influenced by these short videos.

[32] This is based on the assumption that the marginal impact of the videos on viewers is 80% of the effect observed in our experiment, due to repeated viewership, and that participants have only a 20% chance of submitting comments—whether nationalist or empowering—because of real-world challenges, such as finding and navigating the website.

8.2 Long-term Impacts

Do our video interventions, which significantly alter respondents' estimates of self and social intolerance and discourage opposing comments, have any long-term effects? We conducted follow-up surveys with participants in the intervention experiment at 6 and 12 months after the initial study, obtaining 173 and 160 valid responses, respectively. The statistical results suggest that participants largely forget both the content of the provision and the short videos they watched. As shown in Tables D1 and D2, only 32 respondents (9.61%) in the follow-up surveys could clearly recall the specific content of the new provisions in the Public Security Administration Punishment Law. Of those who watched the videos in our baseline experiment, approximately 61 respondents (18.32%) remember the main content of the videos, and only 49 respondents (14.71%) express ongoing concern about developments related to the legal amendments.

Figure 6 and Table D1 display respondents' perceptions of others' tolerance levels across treatment groups. Notably, the belief that expressing opposition is a "traitor" has markedly changed in all groups, generally aligning with the control group, which did not view the video. As shown in Table D2, few respondents in the follow-up survey maintained opinions similar to those in the initial experiment.

The decreasing impact of these videos and the wane of public interest in this legislation is reflected in the decline in related search interest. Analyzing Baidu's search index for the latter half of 2023, Figure 7 shows that in September 2023, when the draft amendment opened for public comment, terms like "Trumped-up charges" and "Public Security Administration Punishment Law" saw a brief surge in attention. However, this interest quickly waned, and there were no further peaks, suggesting a lack of continued public engagement with the legislative process. Similarly, the search term "Nation" followed this pattern, indicating that public interest, even with nationalism sentiments, was short-lived.

These findings suggest that while the initial interventions have a strong impact on respondents' willingness to provide feedback, this effect diminishes over time as the specific content is forgotten. This highlights the fleeting nature of public interest, creating

opportunities for those with vested interests to influence public opinion. In today's short-video landscape, self-interested actors can steer public attention through targeted video campaigns, prompting policy changes. Once the public comment period ends and public opinion loses relevance, the attention around the event naturally fades, likely due to changes in people's attitudes, without long-term repercussions.

9 Conclusion

Using an online intervention experiment with more than 2,000 respondents, this study investigates how short videos influence public attitudes. The experiment is set in the specific context of the Chinese government's revision of the Public Security Administration Punishment Law, where ambiguous clauses like "hurting national feelings" may indirectly broaden administrative authority.

Our results indicate that respondents exposed to nationalist videos were significantly more hesitant in choosing to submit opposing suggestions. A clear spiral of silence effect was observed, where individuals, after viewing nationalism short videos, tended to overestimate the intolerance of the consensus of public opinion for differing viewpoints. This social pressure led to fewer opposing suggestions. This effect was more pronounced in regions with a legacy of political movements or historical events, such as the Cultural Revolution and the enforcement of the one-child policy, which have undermined social trust. Wealthier and more educated individuals also exhibited stronger conformity tendencies. We estimate that during the public consultation period for this legislation, the proliferation of nationalist videos on short video platforms likely suppressed the expression of opposing opinions by nearly 60%, potentially enabling the passage of vague laws that increase administrative authority.

Acknowledgement

This study received no funding. The authors declare that there are no potential conflicts of interest with any relevant parties.

References

Allcott, Hunt, & Rogers, Todd. 2014. The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, **104**(10), 3003–3037.

Askay, David A. 2015. Silence in the crowd: The spiral of silence contributing to the positive bias of opinions in an online review system. *New Media & Society*, **17**(11), 1811–1829.

Bai, Liang, & Wu, Lingwei. 2023. Political conflict and development dynamics: economic legacies of the cultural revolution. *The Journal of Economic History*, **83**(4), 981–1017.

Bursztyn, Leonardo, & Yang, David Y. 2022. Misperceptions about others. *Annual Review of Economics*, **14**(1), 425–452.

Bursztyn, Leonardo, Egorov, Georgy, & Fiorin, Stefano. 2020a. From extreme to mainstream: The erosion of social norms. *American Economic Review*, **110**(11), 3522–3548.

Bursztyn, Leonardo, González, Alessandra L, & Yanagizawa-Drott, David. 2020b. Misperceived social norms: Women working outside the home in Saudi Arabia. *American Economic Review*, **110**(10), 2997–3029.

Bursztyn, Leonardo, Egorov, Georgy, Haaland, Ingar, Rao, Aakaash, & Roth, Christopher. 2023a. Justifying dissent. *The Quarterly Journal of Economics*, **138**(3), 1403–1451.

Bursztyn, Leonardo, Rao, Aakaash, Roth, Christopher, & Yanagizawa-Drott, David. 2023b. Opinions as facts. *The Review of Economic Studies*, **90**(4), 1832–1864.

Cai, Yong, & Feng, Wang. 2021. The social and sociological consequences of China's one-child policy. *Annual Review of Sociology*, **47**(1), 587–606.

Cantoni, Davide, Yang, David Y, Yuchtman, Noam, & Zhang, Y Jane. 2019. Protests as strategic games: experimental evidence from Hong Kong's antiauthoritarian movement. *The Quarterly Journal of Economics*, **134**(2), 1021–1077.

Chen, Yi, Fan, Ziying, Gu, Xiaomin, & Zhou, Li-An. 2020. Arrival of young talent: The send-down movement and rural education in China. *American Economic Review*, **110**(11), 3393–3430.

Chen, Yuyu, & Yang, David Y. 2019. The impact of media censorship: 1984 or brave new world? *American Economic Review*, **109**(6), 2294–2332.

Clemente, Marco, & Roulet, Thomas J. 2015. Public opinion as a source of deinstitutionalization: A "spiral of silence" approach. *Academy of Management Review*, **40**(1), 96–114.

Cohn, Alain, Engelmann, Jan, Fehr, Ernst, & Maréchal, Michel André. 2015. Evidence for countercyclical risk aversion: An experiment with financial professionals. *American Economic Review*, **105**(2), 860–885.

Engesser, Sven, Ernst, Nicole, Esser, Frank, & Büchel, Florin. 2017. Populism and social media: How politicians spread a fragmented ideology. *Information, Communication & Society*, **20**(8), 1109–1126.

Furnas, Alexander C, & LaPira, Timothy M. 2024. The people think what I think: False consensus and unelected elite misperception of public opinion. *American Journal of Political Science*.

Gao, Yichang, Liu, Fengming, & Gao, Lei. 2023. Echo chamber effects on short video platforms. *Scientific Reports*, **13**(1), 6282.

Gearhart, Sherice, & Zhang, Weiwu. 2018. Same spiral, different day? Testing the spiral of silence across issue types. *Communication Research*, **45**(1), 34–54.

Gelfand, Michele J, Gavrilets, Sergey, & Nunn, Nathan. 2024. Norm dynamics: Interdisciplinary perspectives on social norm emergence, persistence, and change. *Annual Review of Psychology*, **75**(1), 341–378.

Gerber, Alan S, & Rogers, Todd. 2009. Descriptive social norms and motivation to vote: Everybody's voting and so should you. *The Journal of Politics*, **71**(1), 178–191.

Gooch, Elizabeth. 2017. Estimating the long-term impact of the Great Chinese Famine (1959–61) on Modern China. *World Development*, **89**, 140–151.

Gunther, Albert C. 1998. The persuasive press inference: Effects of mass media on perceived public opinion. *Communication Research*, **25**(5), 486–504.

Huang, Huiping. 2005. A cross-cultural test of the spiral of silence. *International Journal of Public Opinion Research*, **17**(3), 324–345.

Huang, Yihong, & Ho, Yuen. 2023. Breaking the Spiral of Silence.

Kalwij, Adriaan, Alessie, Rob, Dinkova, Milena, Schonewille, Gea, Van der Schors, Anna, & Van der Werf, Minou. 2019. The effects of financial education on financial literacy and savings behavior: Evidence from a controlled field experiment in Dutch primary schools. *Journal of Consumer Affairs*, **53**(3), 699–730.

Kast, Felipe, Meier, Stephan, & Pomeranz, Dina. 2018. Saving more in groups: Field experimental evidence from Chile. *Journal of Development Economics*, **133**, 275–294.

Laor, Tal. 2024. Breaking the silence: the role of social media in fostering community and challenging the spiral of silence. *Online Information Review*, **48**(4), 710–724.

Li, Lingyao, Johnson, John, Aarhus, William, & Shah, Dhawal. 2022. Key factors in MOOC pedagogy based on NLP sentiment analysis of learner reviews: What makes a hit. *Computers & Education*, **176**, 104354.

Lu, Yingdan, Schaefer, Jack, Park, Kunwoo, Joo, Jungseock, & Pan, Jennifer. 2024. How information flows from the world to China. *The International Journal of Press/Politics*, **29**(2), 305–327.

Luo, Zhifan, & Li, Muyang. 2024. Participatory censorship: How online fandom community facilitates authoritarian rule. *New Media & Society*, **26**(7), 4236–4254.

Matthes, Jörg, Knoll, Johannes, & von Sikorski, Christian. 2018. The "spiral of silence" revisited: A meta-analysis on the relationship between perceptions of opinion support and political opinion expression. *Communication Research*, **45**(1), 3–33.

Meng, Xin, Qian, Nancy, & Yared, Pierre. 2015. The institutional causes of China's great famine, 1959–1961. *The Review of Economic Studies*, **82**(4), 1568–1611.

Miller, Dale T, Monin, Benoit, & Prentice, Deborah A. 1999. Pluralistic ignorance and inconsistency between private attitudes and public behaviors. *Pages 95–113 of: Attitudes, behavior, and social context*. Psychology Press.

Otis, Eileen M. 2011. *Markets and bodies: Women, service work, and the making of inequality in China*. Stanford University Press.

Perez-Truglia, Ricardo, & Cruces, Guillermo. 2017. Partisan interactions: Evidence from a field experiment in the united states. *Journal of Political Economy*, **125**(4), 1208–1243.

Qin, Bei, Strömberg, David, & Wu, Yanhui. 2018. Media bias in China. *American Economic Review*, **108**(9), 2442–2476.

Reisach, Ulrike. 2021. The responsibility of social media in times of societal and political manipulation. *European Journal of Operational Research*, **291**(3), 906–917.

Scheufele, Dietram A, Shanahan, James, & Lee, Eunjung. 2001. Real talk: Manipulating the dependent variable in spiral of silence research. *Communication Research*, **28**(3), 304–324.

Shi, Yaojiang, & Kennedy, John James. 2016. Delayed registration and identifying the "missing girls" in China. *The China Quarterly*, **228**, 1018–1038.

Su, Conghui, Zhou, Hui, Gong, Liangyu, Teng, Binyu, Geng, Fengji, & Hu, Yuzheng. 2021. Viewing personalized video clips recommended by TikTok activates default mode network and ventral tegmental area. *NeuroImage*, **237**, 118136.

Walder, Andrew G, & Su, Yang. 2003. The cultural revolution in the countryside: Scope, timing and human impact. *The China Quarterly*, **173**, 74–99.

Wang, Yuhua. 2021. The political legacy of violence during China's Cultural Revolution. *British Journal of Political Science*, **51**(2), 463–487.

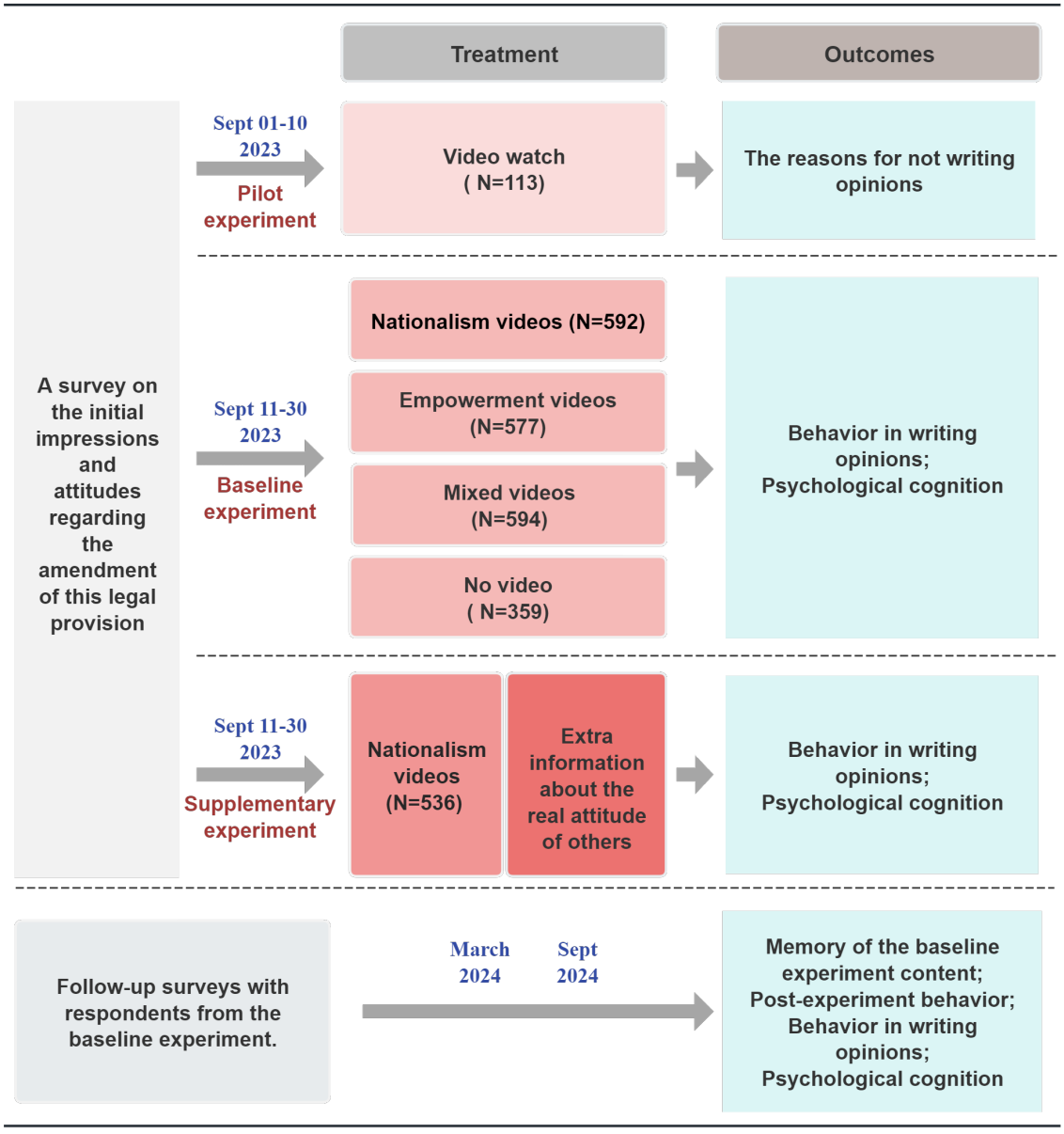
Wang, Yunwen. 2020. Humor and camera view on mobile short-form video apps influence user experience and technology-adoption intent, an example of TikTok (DouYin). *Computers in Human Behavior*, **110**, 106373.

Wojcieszak, Magdalena, & Rojas, Hernando. 2011. Hostile public effect: Communication diversity and the projection of personal opinions onto others. *Journal of Broadcasting & Electronic Media*, **55**(4), 543–562.

Yan, Fei. 2024. "Turning One' s Back on the Party and the People" : Suicides during the Chinese Cultural Revolution. *The China Journal*, **91**(1), 67–88.

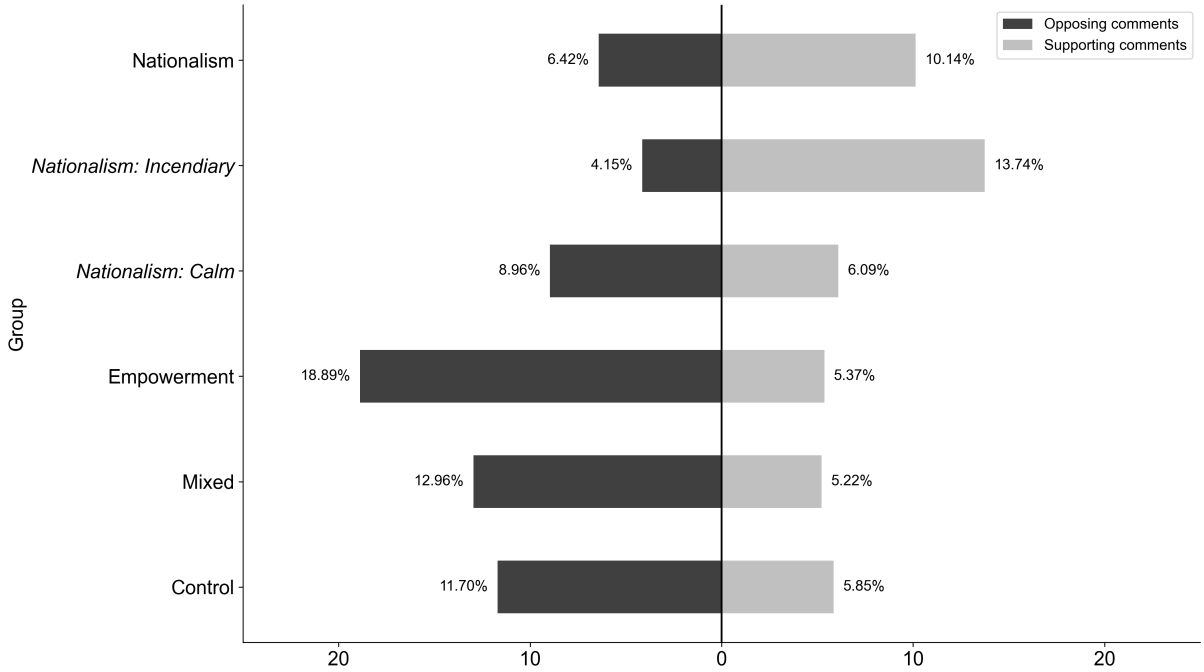
Zhu, Chengyan, Xu, Xiaolin, Zhang, Wei, Chen, Jianmin, & Evans, Richard. 2020. How health communication via Tik Tok makes a difference: A content analysis of Tik Tok accounts run by Chinese provincial health committees. *International Journal of Environmental Research and Public Health*, **17**(1), 192.

FIGURE 1: EXPERIMENTAL FLOW



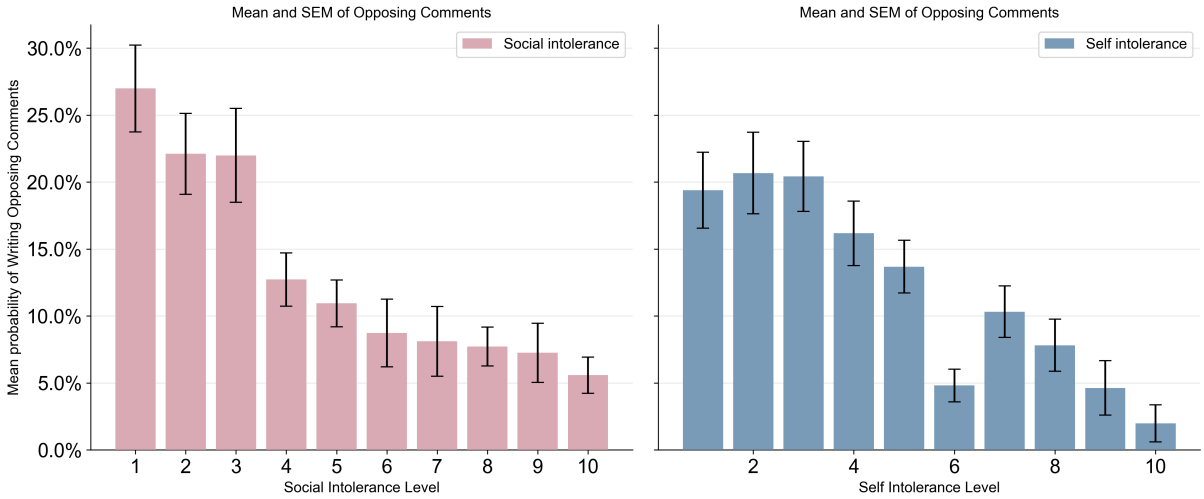
Notes: This figure summarizes the design of the three experiments we conducted. In both the baseline and supplementary experiments, participants watched one-minute videos. Before the baseline experiments, a pilot study was conducted to identify the primary three reasons why respondents were reluctant to express their opinions. In both the baseline and supplementary experiments, participants were given targeted textual information addressing the top three reasons for withholding opinions to rule out the impact of confounding factors. In the supplementary experiment, participants were informed of the actual responses from baseline participants, allowing us to separate the effects of persuasion from the spiral of silence. Follow-up surveys were administered six months and one year after the baseline experiment, repeating the same questions without any video or textual interventions.

Figure 2: The Likelihood of Commenting Across Different Groups



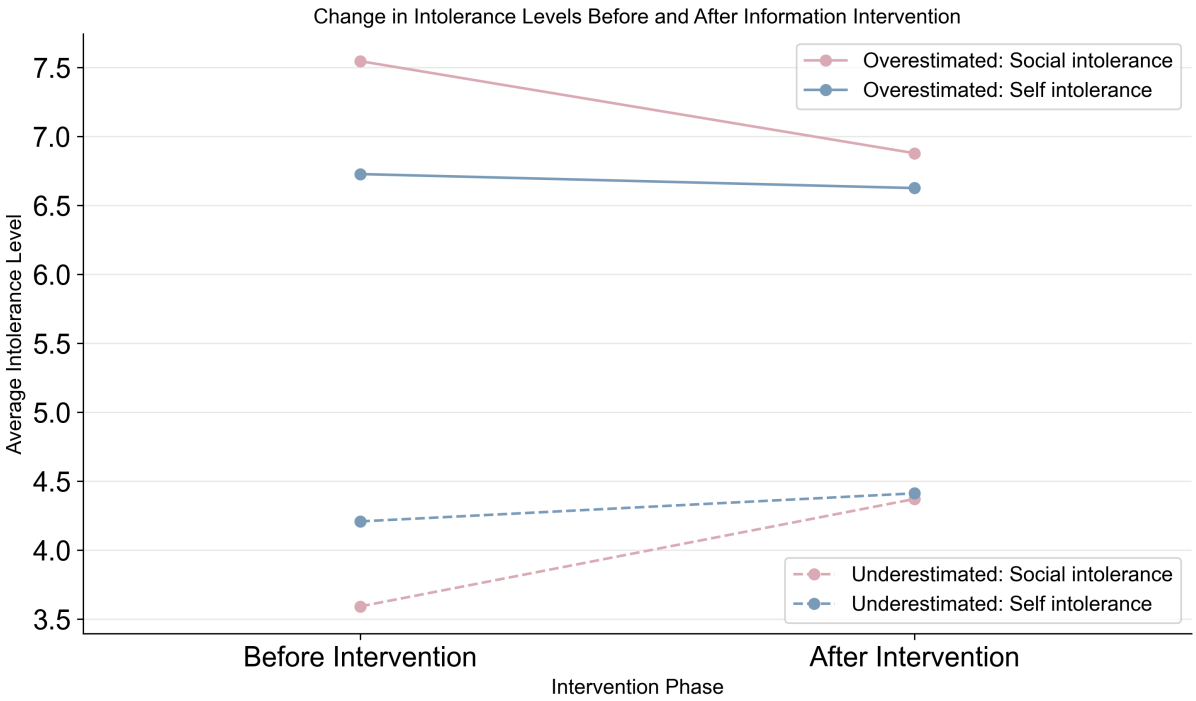
Notes: This figure shows the proportion of the respondents writing down the opposing comments and the supporting comments. The proportion of opposing comments refers to the percentage of each group that wrote down modifications or expressed dissatisfaction with the newly added provisions of the Public Security Administration Punishment Law. The proportion of supporting comments refers to the percentage that wrote down support for this law or criticized those who opposed it.

Figure 3: Average Probability of Respondents Writing opposing Opinions



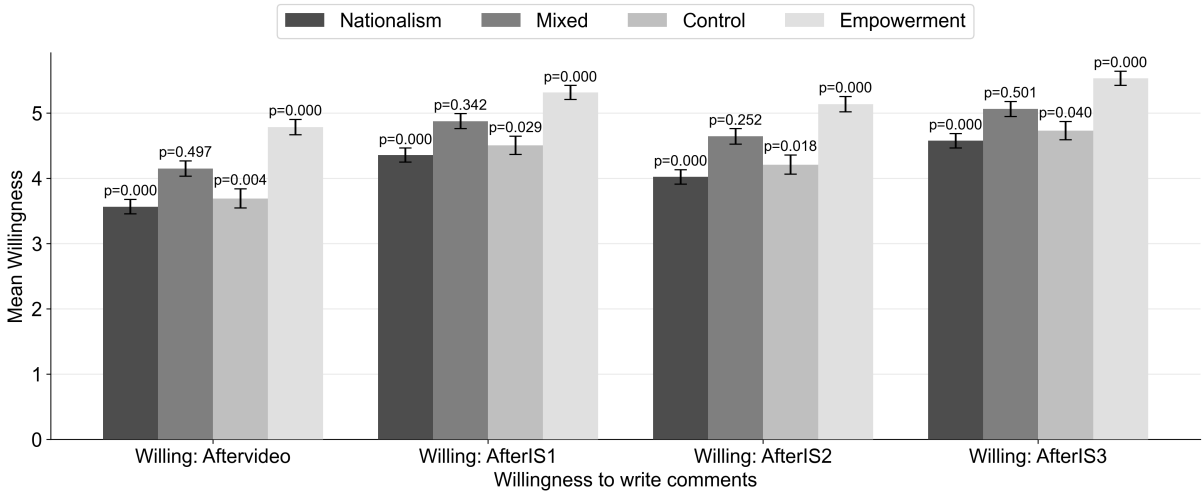
Notes: This figure shows the average likelihood (Mean) and the standard error of the mean (SEM) of respondents expressing opposing views across different levels of self- and social-intolerance. Self-intolerance refers to the respondents' own reluctance to express opposing views, while social intolerance refers to the expectation of others' intolerance toward opposing comments.

Figure 4: Supplementary Experiment: Telling Real Information of Others



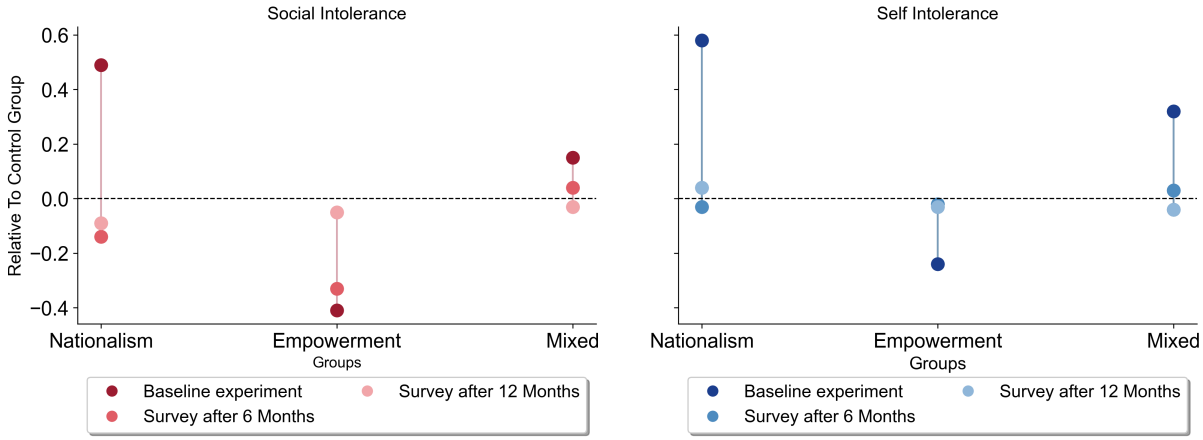
Notes: This figure shows the changes in respondents' perceptions before and after being informed of others' actual levels of intolerance in the supplementary experiment. Self-intolerance refers to the respondents' own reluctance to express opposing views (blue line), while social intolerance refers to the expectation of others' intolerance toward opposing comments (pink line). The solid line represents respondents whose perceived intolerance of others was higher than the actual level before receiving the information, while the dashed line represents those whose perceived intolerance was lower than the actual level.

Figure 5: The Willingness of Respondents at Each Stage



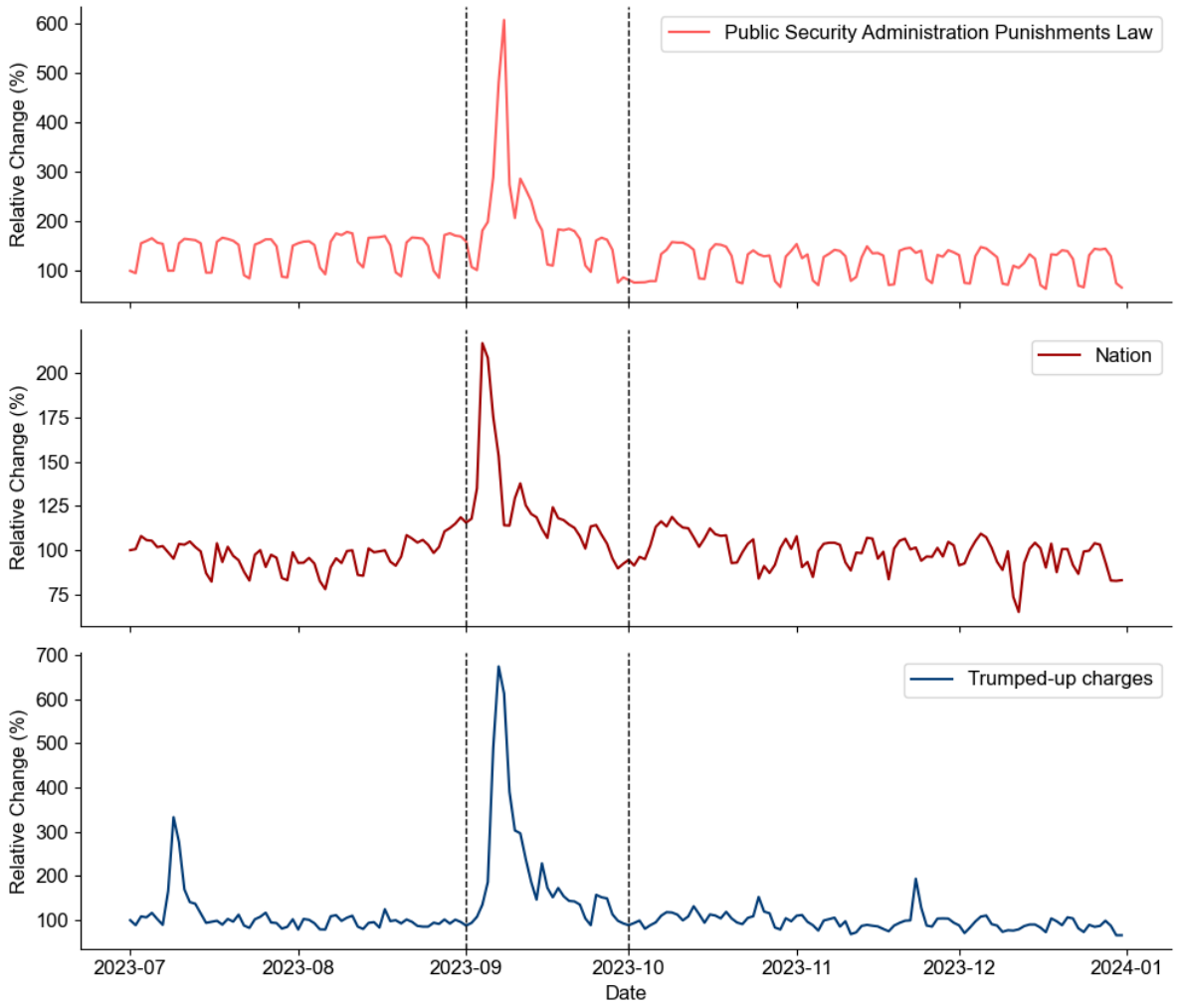
Notes: This figure shows the impact of the self-reported willingness of respondents in each group to propose opposing comments at different stages, including immediately after the video intervention and after each of the three rounds of information shocks. $IS1$ refers to informing respondents about the dangers of power abuse, $IS2$ involves informing them about the public's general enthusiasm for proposing amendments, and $IS3$ provides examples of successful legal changes driven by public initiative. Full details of the content are available in Appendix A1. The P-value indicates whether there is a significant difference between this group and the other three groups.

Figure 6: Changes Between the Basic Experiment and Return Visit Survey



Notes: This figure shows the relative performance in self- and social-intolerance of respondents to the control group during the baseline experiment and the 6-month and 12-month follow-ups. Self-intolerance refers to the respondents' own reluctance to express opposing views, while social intolerance refers to the expectation of others' intolerance toward opposing comments.

Figure 7: Network Attention Trend of Related Keywords



Notes: This figure shows the search volume for keywords related to the law amendment in the second half of 2023. The shaded period in the middle represents the public consultation window, with higher values indicating more frequent searches and greater public attention.

Table 1: Summary of Information by Sub-group

	Overall	Information of each sub-group			
		Control	Nationalism	Empowerment	Mixed
Gender	0.563 (0.496)	0.538 (0.499)	0.576 (0.495)	0.568 (0.496)	0.561 (0.497)
Age	3.403 (1.039)	3.390 (0.996)	3.478 (1.019)	3.435 (1.036)	3.306 (1.082)
Education	2.833 (0.682)	2.836 (0.663)	2.802 (0.686)	2.870 (0.663)	2.827 (0.709)
Income	2.761 (1.067)	2.864 (1.028)	2.802 (1.076)	2.756 (1.055)	2.663**(1.089)
Attitude: Patriotism	1.035(0.611)	1.000(0.598)	1.024(0.612)	1.050(0.615)	1.052(0.614)
Attitude: Cooperation	5.745 (2.654)	5.801 (2.640)	5.594 (2.705)	5.899 (2.548)	5.719 (2.706)
Attitude: Money	7.509 (1.902)	7.474 (2.031)	7.627 (1.786)	7.387 (1.897)	7.526 (1.938)
Attitude: Distrust	6.400 (1.613)	6.322 (1.671)	6.454 (1.578)	6.417 (1.522)	6.375(1.694)
Obs	2122	359	592	577	594

Notes: This table shows the mean and standard deviation (in parentheses) of respondent characteristics for the four main groups. T-tests are conducted to compare the characteristics of respondents in each of the three treatment groups with those in the control group. In our experiments, we tested for respondents' initial levels of and their attitudes toward cooperation, trust, and value of money. The ***, **, and * indicate significant difference from the control group at the significance levels 1%, 5%, and 10%, respectively.

Table 2: Baseline Results: Impact of Video Intervention on Writing Opposing Comments

	(1)	(2)	(3)	(4)	(5)	(6)
D.V.	$\mathbf{1}_{opposing}$		**Emotional tendency**			
			Bag of words		SnowNLP	
$V_{\mathbf{Nat}}$	-0.0643***		-0.0483**		-0.0695***	
	(-3.00)		(-2.51)		(-3.21)	
$V_{Nat:\ Inc}$		-0.0873***		-0.0399*		-0.0866***
		(-2.98)		(-1.66)		(-2.98)
$V_{Nat:\ Calm}$		-0.0425*		-0.0285		-0.0488**
		(-1.88)		(-1.40)		(-2.14)
$V_{\mathbf{Emp}}$	0.0790***	0.0749***	0.0646***	0.0660***	0.0780***	0.0739***
	(4.94)	(4.54)	(4.61)	(4.52)	(4.92)	(4.51)
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Demographics FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs	2,122	2,122	2,122	2,122	2,122	2,122

Notes: This table shows the average marginal effects results of a Logit regression analysis, examining the impact of Nationalism and Empowerment videos on respondents' likelihood to write opposing opinions to the specific legal provision. $V_{\mathbf{Nat}}$ and $V_{\mathbf{Emp}}$ represent the nationalist and empowerment video interventions, respectively. $V_{Nat:\ Inc}$ and $V_{Nat:\ Calm}$ refer to incendiary and calm nationalist video, respectively. The first two columns show the influence on the likelihood of writing opposing opinions, while the latter four columns reflect the emotional tendencies expressed in these opposing opinions. Emotional tendencies are measured using two different methods: The bag of words and the SnowNLP. In the first method, we constructed a sentiment dictionary using a Chinese lexicon, and then combined it with Jieba for word segmentation and a Naive Bayes bag-of-words model to evaluate each review, where a score of 1 indicates negative sentiment (-1 for positive and 0 for neutral sentiments). In the second method, we used the SnowNLP model to directly score the sentiment tendency of each review, with higher scores indicating stronger negative sentiment. The robust z-statistics presented in parentheses. We control for province fixed effects and individuals' demographics, including age, income, gender, education level, and prior attention to the revision. The ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 3: Step One: Outcomes of Video Intervention on Respondents' Self- and Social Intolerance

	(1)	(2)	(3)	(4)
D.V.	**Self intolerance**		**Social intolerance**	
$V_{\mathbf{Nat}}$	0.3631***		0.3912***	
	(2.71)		(2.65)	
$V_{Nat:\ Inc}$		0.7953***		0.8552***
		(4.68)		(4.72)
$V_{Nat:\ Calm}$		0.2544*		0.1961
		(1.69)		(1.16)
$V_{\mathbf{Emp}}$	-0.4272***	-0.3380***	-0.4713***	-0.3972**
	(-3.37)	(-2.58)	(-3.10)	(-2.53)
Province FE	Yes	Yes	Yes	Yes
Demographics FE	Yes	Yes	Yes	Yes
Obs	2,122	2,122	2,122	2,122
R-squared	0.0607	0.0676	0.0689	0.0749
Adjusted R-squared	0.0408	0.0474	0.0491	0.0548

Notes: This table shows the impact of video intervention on respondents' self- and expected social-intolerance. It includes whether the respondents perceive expressing opposition to legal provisions as being labeled a "traitor" (*Self intolerance*), and whether they believe others would perceive it in the same way (*Social intolerance*). $V_{\mathbf{Nat}}$ and $V_{\mathbf{Emp}}$ represent the nationalist and empowerment video interventions, respectively. $V_{Nat:\ Inc}$ and $V_{Nat:\ Calm}$ refer to incendiary and calm nationalist video, respectively. The robust z-statistics presented in parentheses. We control for province fixed effects and individuals' demographics, including age, income, gender, education level, and prior attention to the revision. The ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 4: Step Two: Effect of Respondents' Self- and Social Intolerance on Writing Opposing Comments

	(1)	(2)	(3)
D.V.		$\mathbf{1}_{opposing}$	
Self intolerance	-0.0231***		-0.0154***
	(-7.46)		(-4.37)
Social intolerance		-0.0219***	-0.0166***
		(-7.98)	(-5.45)
Province FE	Yes	Yes	Yes
Demographics FE	Yes	Yes	Yes
Obs	2,122	2,122	2,122

Notes: This table shows the impact of two intolerance levels: the respondents' intolerance (*Self intolerance*) on their behavior of writing opposing comments ($\mathbf{1}_{opposing}$) and their judgment of how others perceive this issue (*Social intolerance*). The robust z-statistics presented in parentheses. We control for province fixed effects and individuals' demographics, including age, income, gender, education level, and prior attention to the revision. The ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 5: Supplementary Experiment: the Effect of Informing Real Intolerance level of Others

	(1)	(2)
D.V.	$\mathbf{1}_{opposing}$	$\mathbf{1}_{supporting}$
PerceptionGap × Informed	0.0830***	0.0020
	(8.76)	(0.02)
Informed	0.1529***	-0.3814***
	(5.99)	(-1.66)
PerceptionGap	-0.0076	-0.1023*
	(-1.03)	(-1.84)
Self intolerance	-0.0361***	0.0744
	(-6.01)	(1.31)
Province FE	Yes	Yes
Demographics FE	Yes	Yes
Obs	1,128	1,128

Notes: This table shows the impact of informing respondents about the true intolerance level of others ($informed$) on their behavior in writing opinions ($\mathbf{1}_{opposing}$ and $\mathbf{1}_{supporting}$). $PerceptionGap$ means the difference between the respondents' expected intolerance of others and the actual value before receiving the information. The robust z-statistics presented in parentheses. We control for province fixed effects and individuals' demographics, including age, income, gender, education level, and prior attention to the revision. The ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 6: Heterogeneity: The Role of Respondent Characteristics in the Impact of Video Interventions

	First Step		Second Step
	(1)	(2)	(3)
D.V.	**Self intolerance**	**Social intolerance**	$\mathbf{1}_{opposing}$
Key Variables			
Culture Revolution	1.2846**	1.9265***	-0.0289**
	(2.38)	(3.51)	(-2.09)
One-child Policy	0.0460***	0.0523***	0.0002
	(2.78)	(2.80)	(0.40)
Famine	0.2737	0.5841**	-0.0037
	(1.12)	(2.12)	(-0.67)
Income	0.2599	0.1481	-0.0164***
	(1.03)	(0.53)	(-2.87)
High Education	-0.0385	-0.0074	-0.0078*
	(-0.27)	(-0.04)	(-1.87)
Age	-0.1255	-0.2701**	-0.0028
	(-1.15)	(-2.39)	(-0.97)
Province FE	Yes	Yes	Yes
Demographics FE	Yes	Yes	Yes
Obs	2,122	2,122	2,122

Notes: This table shows the heterogeneity effects of various respondent characteristics at each of the two steps of the spiral of silence. In the first step, we examine which characteristics make respondents more susceptible to nationalist videos interference, leading to changes in their self-intolerance and predictions of others' intolerance. In the second step, we examine which characteristics make respondents more likely to avoid expressing opposing opinions when facing pressure from others (measured by the *PerceptionGap*, which defined in Section 5). The first two columns show the average effects of the interaction term $V_{Nat} \times KeyVariables$, while column (3) presents those of the term $PerceptionGap \times KeyVariables$. The robust z-statistics presented in parentheses. We control for province fixed effects and individuals' demographics, including age, income, gender, education level, and prior attention to the revision. The ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 7: The Spread of Video Content and its Influence on Commenting

Panel A: Total Video Views (Thousands)

		2023.8.1-2023.8.31		2023.9.1-2023.9.30		2023.10.1-2023.10.30	
	Online Platform	**Tiktok**	**Kuaishou**	**Tiktok**	**Kuaishou**	**Tiktok**	**Kuaishou**
Video Type	Nationalism	0	0	19,380	1,860	560	280
	Empowerment	0	0	2,140	200	220	60

Panel B: Reduction in the Number of Opposing Comments Due to Short Video Exposure

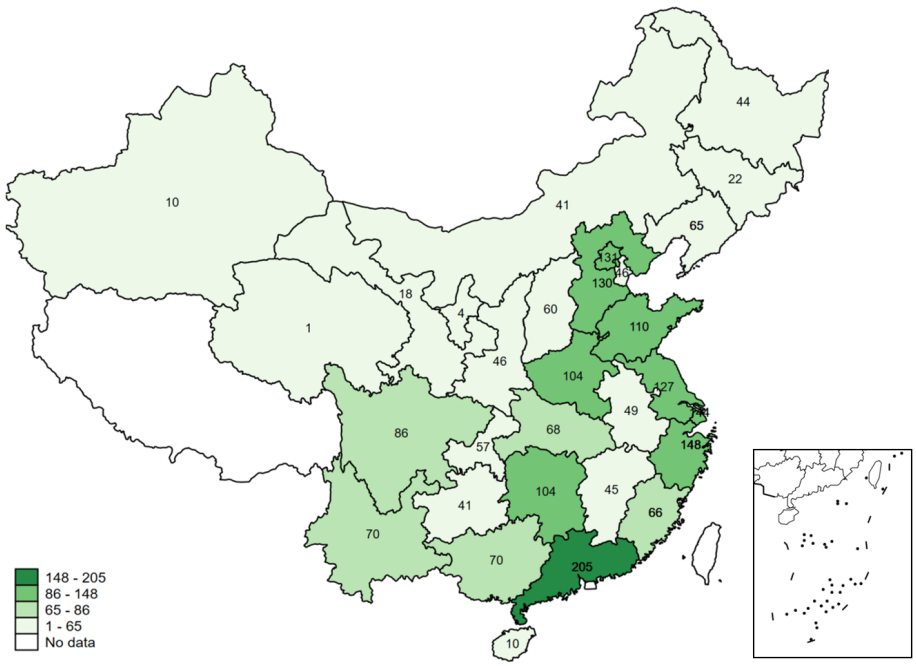
		Discount Factor: Real-World Difficulties		
		30%	20%	10%
Discount Factor:	90%	318,913	212,609	106,304
Repeated	80%	283,478	188,985	94,493
Viewers	70%	248,943	165,362	82,681

Notes: This table shows the total views and impacts of the short videos identified using the method in Figure D1. Panel A provides an estimate of the views for videos discussing the legal amendment during the public consultation period on China's top two short video platforms, TikTok and Kuaishou. Since the total view count is unavailable, we use the total number of "likes" and apply a conservative 5% like-to-view ratio to estimate the total views. Panel B estimates the reduction in opposing comments in real-world scenarios, based on our experimental results and accounting for other factors that might influence individuals' willingness to comment. The analysis includes two discount factors: one that reflects real-world difficulties in commenting (Real-World Commenting Difficulties), such as challenges in finding or using the website, and another that accounts for the declining marginal impact of videos on viewers who watch them several times (Repeated Viewers).

Appendix

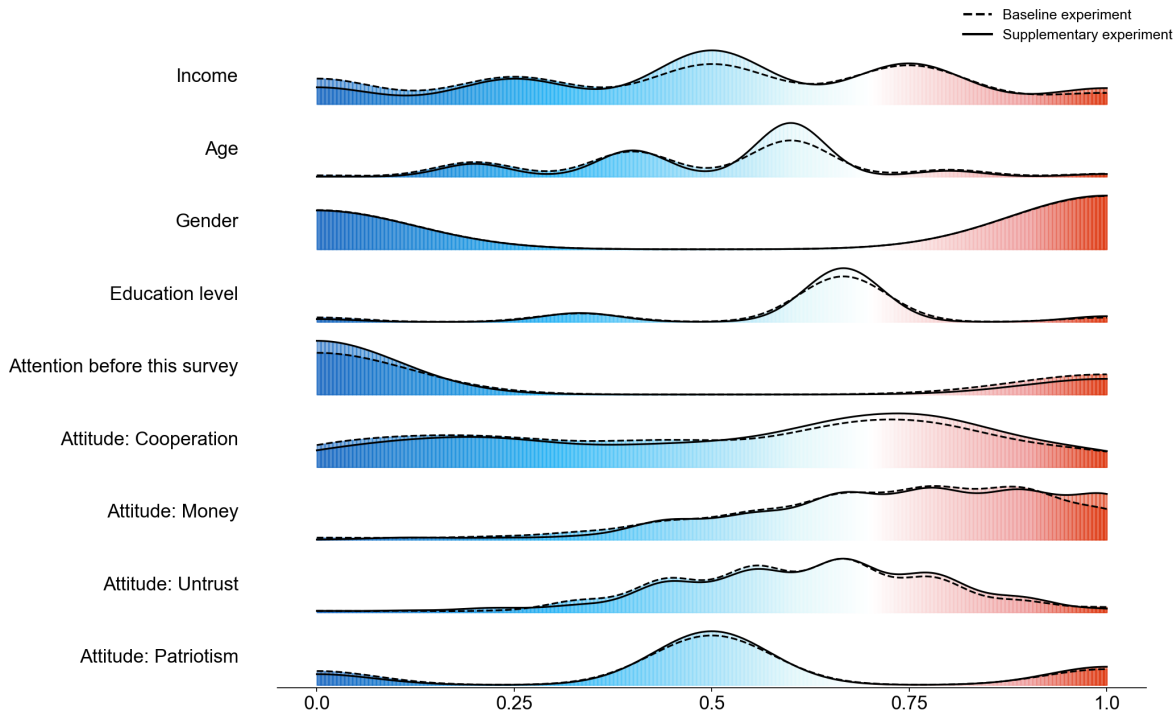
A Supplementary Sample Information

Figure A1: Distribution of Participants



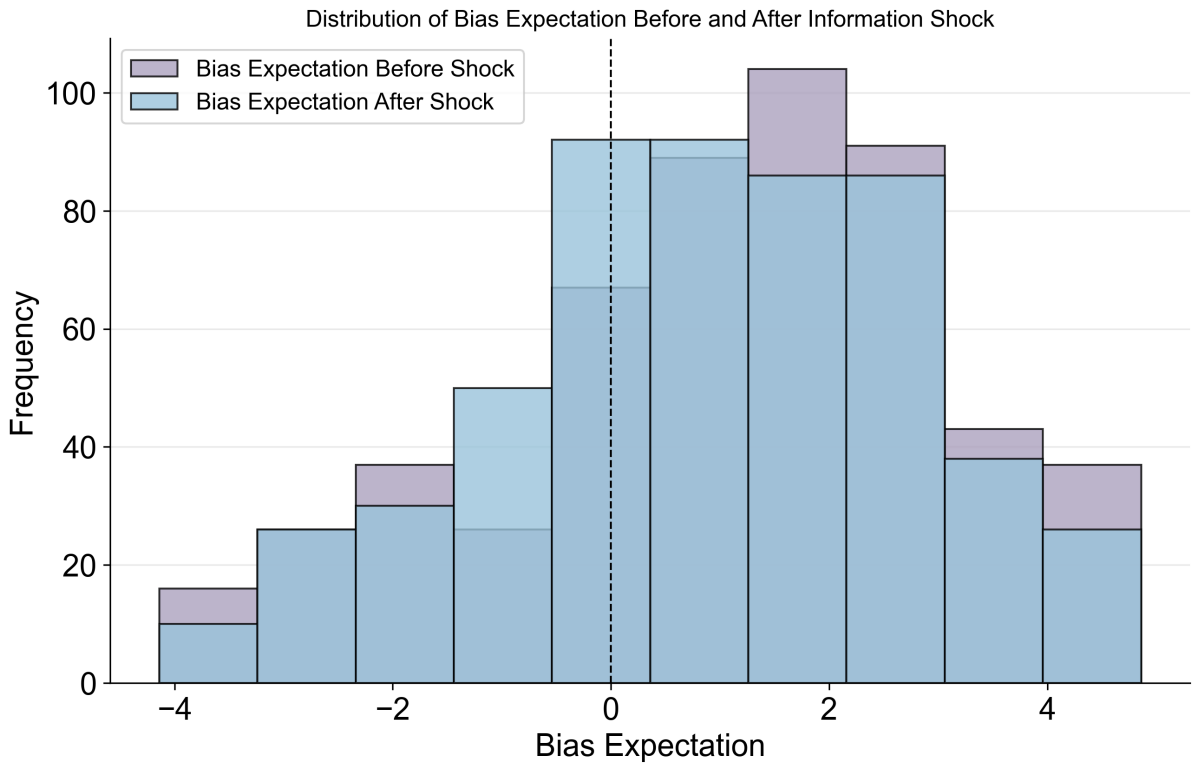
Notes: This figure shows the geographical distribution of our participants.

Figure A2: Sample Distribution of The Nationalism Group In Two Experiments



Notes: This figure shows the distribution of the respondents of the nationalism video groups in the baseline and supplementary experiments. We compare the differences in sample distribution based on basic characteristics such as gender, as well as attitudes towards nationalism, money, effort, distrust, and whether they had paid attention to the legal amendment prior to participating in the survey. The bandwidth for this kernel density estimation is 0.25.

Figure A3: Supplementary Experiment: Telling Real Information of Others



Notes: This figure demonstrates that the changes in the bias of the respondents in our supplementary experiment before and after receiving information shock (the average intolerance level of others). The vertical dashed line represents the average intolerance level of others that we informed respondents.

Table A1: The Content of Video Interventions and Information Shocks

Panel A: The basic content of video intervention		
Type	**Video piece**	**Theme**
Nationalism	Incendiary Nationalism	Use extreme and inflammatory lines and pictures to link this law revision with safeguarding national dignity.
	Calm Nationalism	Use objective lines and descriptions to explain the significance of this legal revision to national feelings.
Empowerment		Linking this legal revision with the expansion and abuse of public power, explaining that this clause is neither clear nor well-regulated.
Panel B: The basic content of three information shocks		
Information shocks	**Theme**	**Main Content**
IS1	The abuse of public power is closely related to everyone.	On December 13, 2014, during the handling of a police incident at a construction site, Officer Wang Wenjun of the Taiyuan Public Security Bureau abused his power, resulting in the wrongful death of migrant worker Zhou Xiuyun from Henan Province.
IS2	The actual number of people who filled out the feedback and the total number of comments.	As of 4:00 PM on September 16, 2023, a total of 93,724 people have submitted their comments on the revision of this regulation on the official website, with a total of 116,463 comments, and the numbers are still increasing.
IS3	A precedent where public feedback has successfully changed the provisions of a legal revision.	The draft revision of the "Law of the People's Republic of China on the Protection of Women's Rights and Interests" was opened for public comment on December 24, 2021. Over 80,000 people submitted 420,000 comments, successfully leading to the removal of improperly or unreasonably worded provisions in the draft.

Table A2: Examples Of Keywords Used In Bag Of Words Analysis

	Negative	Positive
Examples of words	主观判断 (biased judgment)	合情 (reasonable)
	选择性执法 (arbitrary enforcement)	合理 (rational / reasonable)
	扣帽子 (name-calling)	合适 (suitable / appropriate)
	贴标签 (stereotyping)	很好 (very good)
	上纲上线 (overgeneralizing)	非常好 (excellent)
	干涉 (intrusion)	很有意义 (meaningful)
	不分黑白 (simplistic thinking)	适合使用 (suitable for use)
	冤枉 (wronged / unjustly accused)	完全合理 (completely reasonable)
	误导 (deceptive)	特别适合 (particularly suitable)
	模棱两可 (equivocal)	完全支持 (fully support)
	泛泛而谈 (talking in platitudes)	应该如此 (it should be so / it is as it should be)
	误判 (misinterpretation)	彰显情怀 (demonstrate sentiment)
	滥用 (abuse)	保证权益 (ensure rights)
	似是而非 (misleadingly plausible)	无异议 (uncontested / without objection)
	乱套 (chaos / disorder)	挺好的 (pretty good / quite good)

Notes: This table displays a portion of the vocabulary used in the bag-of-words method for analyzing the sentiment tendencies of the comments submitted by respondents. Our vocabulary is based on the NTUSD Chinese Sentiment Dictionary published by National Taiwan University, and it incorporates certain keywords or expressions closely related to this legal revision. Positive words refer to those that reflect sentiments expressed by respondents in a commendatory or affirmative manner. Negative words refer to those that reflect sentiments expressed by respondents in a derogatory or negative manner. We combine word segmentation processing to segment the respondents' comments and determine their sentiment tendencies.

Table A3: Correlation of Respondents' Self-reported Willingness With Writing the Revised Opinion

	$\mathbf{1}_{opposing}$	Willing: Aftervideo	Willing: AfterIS1	Willing: AfterIS2	Willing: AfterIS3
$\mathbf{1}_{opposing}$	1.0000				
Willing: Aftervideo	0.2773***	1.0000			
Willing: AfterIS1	0.2609***	0.6665***	1.0000		
Willing: AfterIS2	0.3230***	0.6750***	0.7371***	1.0000	
Willing: AfterIS3	0.3413***	0.5901***	0.6922***	0.7967***	1.0000

Notes: This table illustrates the correlation between the self-reported scores of respondents' willingness to voice objections at various stages and their actual behavior of writing down objections. It can be observed that the respondents' self-reported willingness is significantly positively correlated with their final behavior, and there is also a significant positive correlation between the willingness at different stages (i.e., the consistency of willingness). Therefore, it can be concluded that the self-reported willingness is reliable, and the experimental process is credible.

B Supplementary Empirical Results

B.1 Empirical Results For Supporting Comments

Table B1: Baseline Results: Impact of Video Intervention on Writing Supporting Comments

D.V.	(1)	(2)	(3)	(4)	(5)	(6)
	$\mathbf{1}_{supporting}$		**Emotional tendency**			
			Bag of words		SnowNLP	
$V_{\mathbf{Nat}}$	0.0516***		0.0416***		0.0522***	
	(4.06)		(3.55)		(4.05)	
$V_{Nat:\ Inc}$		0.0605***		0.0388***		0.0597***
		(3.75)		(2.63)		(3.63)
$V_{Nat:\ Calm}$		0.0551***		0.0460***		0.0572***
		(3.71)		(3.43)		(3.80)
$V_{\mathbf{Emp}}$	0.0071	0.0137	0.0068	0.0115	0.0048	0.0120
	(0.35)	(0.86)	(0.52)	(0.84)	(0.32)	(0.85)
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Demographics FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs	2,122	2,122	2,122	2,122	2,122	2,122

Notes: This table presents the results of a Logit regression analysis, examining the impact of Nationalism and Empowerment videos on respondents' likelihood to write supporting opinions (i.e., expressing support to the specific legal provision). $V_{\mathbf{Nat}}$ and $V_{\mathbf{Emp}}$ represent the nationalist and empowerment video interventions, respectively. $V_{Nat:\ Inc}$ and $V_{Nat:\ Calm}$ refer to incendiary and calm nationalist video, respectively. Emotional tendencies are measured using two different methods: The bag of words and the SnowNLP. In Method 1, we constructed a sentiment dictionary using a Chinese lexicon, and then combined it with Jieba for word segmentation and a Naive Bayes bag-of-words model to evaluate each review, where a score of 1 indicates negative sentiment. In Method 2, we used the SnowNLP model to directly score the sentiment tendency of each review, with higher scores indicating stronger negative sentiment. We manually reviewed the output results for each comment. The robust z-statistics presented in parentheses. Fixed effects of province and individuals' demographics (including age, income, gender, education level, and attention to this revision before participating in our experiment) are included in the analysis. The ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table B2: Step Two: Effect of Respondents' Cognition on Writing Supporting Comments

	(1)	(2)	(3)
D.V.		$\mathbf{1}_{supporting}$	
Self intolerance	0.0063***		0.0076***
	(2.82)		(2.97)
Social intolerance		0.0005	-0.0026
		(0.26)	(-1.22)
Province FE	Yes	Yes	Yes
Demographics FE	Yes	Yes	Yes
Obs	2,122	2,122	2,122

Notes: This table shows the impact of two types of cognition on respondents' behavior in writing supporting comments. It includes both the respondents' intolerance of expressing opposition as "western obsession" (*Self intolerance*) and their judgment of how others perceive this issue (*Social intolerance*) affect their writing behavior ($\mathbf{1}_{supporting}$). $V_{\mathbf{Nat}}$ and $V_{\mathbf{Emp}}$ represent the nationalist and empowerment video interventions, respectively. $V_{Nat:\ Inc}$ and $V_{Nat:\ Calm}$ refer to incendiary and calm nationalist video, respectively. The robust z-statistics presented in parentheses. Fixed effects of province and individuals' demographics (including age, income, gender, education level, and attention to this revision before participating our experiment) are included in the analysis. The ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

B.2 Channel Exclusion

Table B3: Channel Exclusion: the Power Distance to the Government

	(1)	(2)	(3)	(4)
D.V.	$\mathbf{1}_{opposing}$		$\mathbf{1}_{supporting}$	
$V_{\mathbf{Nat}}$	-0.0584***		0.0526***	
	(-2.78)		(4.09)	
$V_{Nat:\ Inc}$		-0.0840***		0.0611***
		(-2.91)		(3.74)
$V_{Nat:\ Calm}$		-0.0369*		0.0567***
		(-1.66)		(3.73)
$V_{\mathbf{Emp}}$	0.0711***	0.0669***	0.0029	0.0116
	(4.52)	(4.12)	(0.20)	(0.75)
Province FE	Yes	Yes	Yes	Yes
Demographics FE	Yes	Yes	Yes	Yes
Power distance FE	Yes	Yes	Yes	Yes
Obs	2,212	2,212	2,212	2,212

Notes: This table presents the empirical results after fixing the effects of respondents' "power distance." Power distance reflects the extent to which respondents fear the government or authority. During the survey process, we asked respondents to rate on a scale of 1 to 10, "How likely do you think it is that the government will adopt your suggestions?" A higher score indicates that the respondents believe the government is more likely to adopt their suggestions, thus representing a shorter power distance. $V_{\mathbf{Nat}}$ and $V_{\mathbf{Emp}}$ represent the nationalist and empowerment video interventions, respectively. $V_{Nat:\ Inc}$ and $V_{Nat:\ Calm}$ refer to incendiary and calm nationalist video, respectively. The robust z-statistics presented in parentheses. Fixed effects of province and individuals' demographics (including age, income, gender, education level, and attention to this revision before participating our experiment) are included in the analysis. The ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table B4: Channel Exclusion: A Complete Projection of Self-perception

Panel A: Correlation of self- and social- intolerance levels		
D.V.	**Self intolerance**	**Social intolerance**
Self intolerance	1.000	
Social intolerance	0.4447***	1.000

Panel B: Results after controlling self-intolerance				
D.V.	$\mathbf{1}_{opposing}$		$\mathbf{1}_{supporting}$	
$V_{\mathbf{Nat}}$	-0.0570***		0.0497***	
	(-2.74)		(3.86)	
$V_{Nat:\ Inc}$		-0.0705**		0.0564***
		(-2.49)		(3.39)
$V_{Nat:\ Calm}$		-0.0372*		0.0539***
		(-1.69)		(3.61)
$V_{\mathbf{Emp}}$	0.0710***	0.0691***	0.0079	0.0157
	(4.48)	(4.23)	(0.56)	(1.04)
Province FE	Yes	Yes	Yes	Yes
Demographics FE	Yes	Yes	Yes	Yes
Self intolerance FE	Yes	Yes	Yes	Yes
Obs	2,212	2,212	2,212	2,212

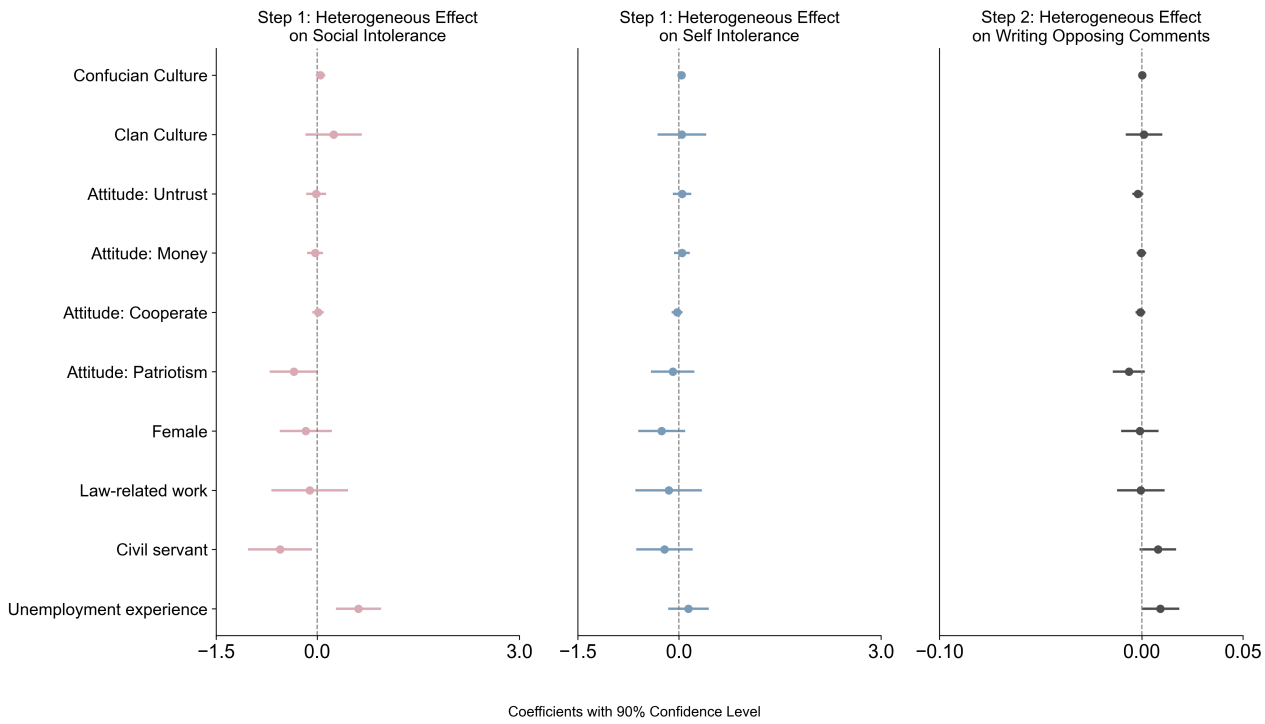
Notes: Panel A of this table shows a positive correlation between respondents' own intolerance and their perceived intolerance of others. If the correlation coefficient is far less than 1, the judgments about others cannot be fully attributed to a projection of one's own consciousness. Panel B presents the empirical results after controlling for the respondents' own opinions. $V_{\mathbf{Nat}}$ and $V_{\mathbf{Emp}}$ represent the nationalist and empowerment video interventions, respectively. $V_{Nat:\ Inc}$ and $V_{Nat:\ Calm}$ refer to incendiary and calm nationalist video, respectively. The robust z-statistics presented in parentheses. Fixed effects of province and individuals' demographics (including age, income, gender, education level, and attention to this revision before participating our experiment) are included in the analysis. The ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

B.3 Heterogeneity Analysis For Other Characteristics

In this section, we present the regression coefficients for the average heterogeneous effects of various respondent characteristics. The first category includes socio-cultural factors. We have highlighted the significant impacts of historical events such as the Cultural Revolution and famine on the spiral of silence. Here, we further explore the influence of clan culture and Confucian culture, which are also highly influential factors. While the main text has already addressed the effects of age, income, and education, in this section, we assess the influence of additional personal traits, including gender. We surveyed respondents on various attitudes, including their trust in others within society, the importance they place on effort in improving quality of life, the significance of money, and their level of nationalism prior to the video intervention. Additionally, we examined the impact of respondents' work characteristics, such as whether they are employed in the legal profession, are civil servants, or have experienced unemployment.

According to the results shown in Figure B1, most of these characteristics do not significantly affect the influence of the video intervention on the behavior of writing opposing comments. Even when there is an effect, it tends to be relatively minor. This further illustrates that, for the spiral of silence effect, historical events that undermine social trust and the concerns of elite groups have the most critical impact.

FIGURE B1: HETEROGENEOUS EFFECTS OF EACH FEATURE IN EACH STEP



Notes: This figure intuitively displays the regression coefficients of the average heterogeneous effects of various respondent characteristics at each step. In the first step, a positive coefficient indicates that the characteristic significantly enhances the impact of the video intervention on psychological cognition. In the second step, a negative coefficient indicates that the characteristic significantly amplifies the influence of pressure from others on the likelihood of writing opposing opinions.

C Robustness tests

C.1 Excluding observations With Ambiguous Attitudes

To strengthen the credibility of our findings, we apply a rigorous filtering process to respondents' comments. We begin by excluding overly simplistic or potentially insincere responses, such as "support" "good" or "needs revision" , which may also have been written for perfunctory reasons.

Additionally, we remove extreme opinions, such as "renounce Chinese citizenship," which may indicate frustration or sarcasm rather than genuine sentiment. After removing, we get a refined sample of 1,933 respondents. Logistic regression analysis conducted on this refined sample, as shown in Table C1, demonstrates that while the average marginal effects coefficients are slightly reduced, the overall results remain consistent with our baseline findings. This process further confirms the robustness of the significant impact of the video intervention.

Table C1: Robustness test: Removing Samples With Ambiguous Comments

	(1)	(2)	(3)	(4)
D.V.	$\mathbf{1}_{opposing}$		$\mathbf{1}_{supporting}$	
$V_{\mathbf{Nat}}$	-0.0507**		0.0542***	
	(-2.41)		(4.16)	
$V_{Nat:\ Inc}$		-0.0855***		0.0632***
		(-2.90)		(3.81)
$V_{Nat:\ Calm}$		-0.0309		0.0596***
		(-1.39)		(3.92)
$V_{\mathbf{Emp}}$	0.0756***	0.0708***	0.0008	0.0102
	(4.71)	(4.28)	(0.05)	(0.62)
Province FE	Yes	Yes	Yes	Yes
Demographics FE	Yes	Yes	Yes	Yes
Obs	1,933	1,933	1,933	1,933

Notes: This table presents the empirical results after removing samples with ambiguous attitudes. In the written opinions, some respondents provided very few words (fewer than 5), and some opinions may have been sarcastic (e.g., "suggest punishing the entire family"). Therefore, we excluded these samples with ambiguous attitudes. $V_{\mathbf{Nat}}$ and $V_{\mathbf{Emp}}$ represent the nationalist and empowerment video interventions, respectively. $V_{Nat:\ Inc}$ and $V_{Nat:\ Calm}$ refer to incendiary and calm nationalist video, respectively. The robust z-statistics presented in parentheses. Fixed effects of province and individuals' demographics (including age, income, gender, education level, and attention to this revision before participating our experiment) are included in the analysis. The ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

C.2 Another Measurement of Social Intolerance

To further substantiate our findings, we employ an alternative approach to assess respondents' perceived external pressure. During the survey, we ask respondents to estimate the proportion of societal support, opposition, and neutrality towards the new regulation (using a scale of 1 to 100). We use a slider setting on the survey platform to allow respondents to allocate proportions among these three groups, making it easy for them to select their answers and ensuring that the total sums to 100%.

- *How many people do you think support this new regulation?*
- *How many do you think oppose it?*
- *How many do you think remain neutral?*

We construct the variable $Social_intoleranceII$, defined as the difference between the perceived proportion of support and opposition, to quantify external pressure. The rationale is that when respondents believe the support proportion significantly outweighs opposition, the pressure to voice a opposing opinion is heightened. The empirical results in Table C2 demonstrate that in the first stage, the nationalist video intervention significantly increased respondents' perceived support-opposition difference by approximately 5.5 points. In the second stage, this increased difference notably reduced the likelihood of respondents writing revised opinions. Specifically, for each 1-unit increase in the difference, the probability of writing a revised opinion decreased by 0.14%.

This alternative measure of external pressure further corroborates the finding that respondents, influenced by the video intervention, perceive greater societal pressure, which leads to a decrease in expressing dissenting opinions. Thus, the robustness of the spiral of silence effect is reaffirmed.

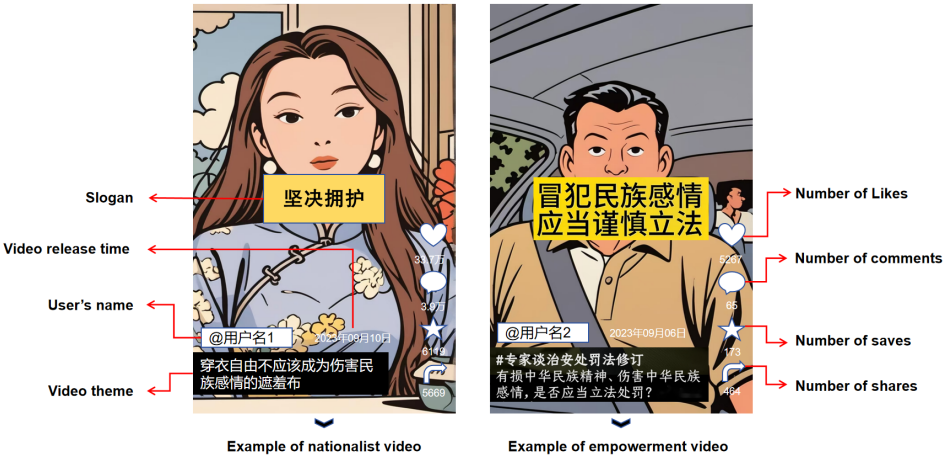
Table C2: Robust test II: Another Measurement of Social Intolerance

	First step	**Second step**
D.V.	Social intoleranceII	$\mathbf{1}_{opposing}$
V_{Nat}	5.4596**	
	(2.57)	
V_{Emp}	-6.7737***	
	(-2.88)	
Social intoleranceII		-0.0014***
		(-10.23)
Province FE	Yes	Yes
Demographics FE	Yes	Yes
Obs	2,122	2,122

Notes: This table presents the empirical results using an alternative method to measure the external intolerance perceived by respondents. In the channel analysis of the baseline regression, we demonstrated in two steps that video interventions influence respondents' self-beliefs and their judgments of others' beliefs, creating pressure that leads to fewer objections being voiced. In this table, we use a different method to measure social intolerance level, specifically by calculating the difference between the proportion of people in society, as perceived by the respondents, who support the policy and those who oppose it. $V_{\mathbf{Nat}}$ and $V_{\mathbf{Emp}}$ represent the nationalist and empowerment video interventions, respectively. $V_{Nat:\ Inc}$ and $V_{Nat:\ Calm}$ refer to incendiary and calm nationalist video, respectively. The robust z-statistics presented in parentheses. Fixed effects of province and individuals' demographics (including age, income, gender, education level, and attention to this revision before participating our experiment) are included in the analysis. The ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

D More Information Of Short- And Long-term Impacts

Figure D1: Example Of Identifying Two Types Of Videos



Notes: This figure shows representative videos on short video platforms related to nationalism and empowerment concerning the revision of this law. We identify whether a video belongs to nationalism or empowerment by searching for keywords related to the legal revision event and analyzing information from video covers and dialogue content.

Table D1: Traced Respondents' psychological performance I

Group	Self intolerance			Social intolerance		
	Baseline	Survey after 6 months	Survey after 12 months	Baseline	Survey after 6 months	Survey after 12 months
Nationalism	5.53	4.06	3.83	6.11	4.58	4.05
Empowerment	4.71	4.07	3.76	5.21	4.39	4.09
Mixed	5.27	4.12	3.75	5.77	4.76	4.11
Control	4.95	4.09	3.79	5.62	4.72	4.14

Notes: This table shows the psychological changes observed in respondents after 6 months and 12 months of follow-up.

Table D2: Traced Respondents' Psychological Performance II

Panel A: The basic impression of the law in the return visit

Group	Remember exactly what the new law is about	Remember the video content	Give comments again	The opinion made is similar to the first opinion
Nationalism	11.49%	19.54%	16.09%	21.43%
Empowerment	3.33%	16.67%	16.67%	33.33%
Mixed	13.64%	32.95%	15.91%	7.14%
Control	10.29%	-	17.65%	25.00%

Panel B: Changes in Respondents' Perception of Overall Social Attitudes

Group	The average proportion of the public that the respondents believe support this regulation (%)		The average difference between support and opposition as perceived by the respondents (%)	
	Before	Later	Before	Later
Nationalism	74.08	66.6	12.94	18.05
Empowerment	61.67	63.4	20.46	19.01
Mixed	65.01	67.58	13.01	14.92
Control	69.15	70.28	13.40	14.24

Notes: The table presents the overall performance of respondents during follow-up surveys at 6 and 12 months. Panel A shows information on respondents' memory of the video content, the legal provisions, and whether they wrote down their opinions again. Panel B displays changes in respondents' perceptions of overall external attitudes compared to the baseline experiment.