2025

HARVARD-YENCHING

INSTITUTE WORKING

PAPER SERIES

WEDGING AND CHILLING: HOW GOVERNMENT-LED RUMOR REBUTTALS HALT ONLINE ACTIVISM WITHOUT PERSUASION

Kaiping Zhang | Tsinghua University
Siyuan Gao | Tsinghua University

Wedging and Chilling: How Government-Led Rumor Rebuttals Halt Online Activism without Persuasion

Authors

Siyuan Gao, Ph.D Candidate, School of Public Policy and Management, Tsinghua University, Beijing, China.

Kaiping Zhang, Ph.D., Associate Professor, Department of Political Science, School of Social Sciences, Tsinghua University, Beijing, China. Harvard-Yenching Institute, Cambridge, Massachusetts, USA. Correspondence regarding this manuscript should be addressed to kaipingz@tsinghua.edu.cn.

Abstract

This study investigates how government-led rumor rebuttals—in contrast to more familiar strategies of censorship or overt persuasion—shape online activism under authoritarian rule. Moving beyond traditional fact-driven or motivated reasoning frameworks, we introduce a theoretical lens focused on "wedging" and "chilling" effects. Instead of changing factual perceptions or core attitudes, state rebuttals operate by fragmenting dissident networks, undermining solidarity, and creating a dominant opinion climate. Using a natural experiment drawn from an incident of online activism in China, we combine a regression discontinuity design with computational text analysis to show that rumor rebuttals significantly reduce information-sharing behaviors, particularly among dissenting voices, while leaving beliefs and factual understandings largely intact. These findings highlight a subtle yet potent form of digital repression: by discouraging visible engagement and weakening collective resistance, the state can restrain online activism without resorting to direct coercion. Our research underscores the dual vulnerability of digital mobilization—while digital tools facilitate rapid mobilization, they remain susceptible to state-driven communicative interventions.

Keywords: Online activism, misinformation, government rumor rebuttal, computational methods, authoritarian politics

1. Introduction

Social media platforms have emerged as central arenas for the dissemination of both information and misinformation, often catalyzing online collective actions(Earl et al., 2022; Pan & Siegel, 2020; Steinert-Threlkeld et al., 2015; Tufekci & Wilson, 2012; Yang, 2009a, 2009b, 2018). In response, authoritarian governments worldwide emoloy various strategies to intervene in these online protests and maintain control over public narratives. These measures range from denying "fake news", redirecting public attention (Munger et al., 2019; Stukal et al., 2017; Woolley & Howard, 2016), co-opting certain citizens to more extreme actions such as censorship and Internet shutdowns (Hobbs & Roberts, 2018; Jansen & Martin, 2015; King et al., 2017; Roberts, 2018; Stukal et al., 2017). Among these tools, government-led rumor rebuttals—official corrections of misinformation—have emerged as a cost-effective mechanism. This approach involves debunking misinformation by providing evidence through official statements, press releases, and other channels. By doing so, governments aim to clarify facts, restore public trust, and prevent large-scale offline actions.

However, a critical question persists: **how do official rebuttals shape online discussions and if at all, suppress activism?** Existing research offers competing theories about the effects of government-led rumor rebuttal. One stream of literature indicates that countering misinformation with rebuttals can effectively correct public misperceptions (Aird et al., 2018; Clayton et al., 2020). This perspective posits that corrections can persuade individuals or, at the very least, inform them about alternative evidence, potentially leading to behavioral changes. Conversely, an opposing viewpoint argues that individuals' preexisting beliefs shape their perceptions more than the content of the rebuttal (Webster & Abramowitz, 2017). Those

with firm ideological commitments may cling to their misperceptions about contentious issues, regardless of the factual accuracy of the information provided. Rather than being swayed by the content of the rebuttal, people interpret corrections through the lens of their preexisting beliefs, resulting in resistance or outright rejection of the government's narrative (Nyhan & Reifler, 2010).

In this study, we propose a new mechanism to explain how government-led rumor rebuttals influence online activism — not through persuasion, but by fostering divisions. Rather than correcting false beliefs or altering political attitudes, we argue that rebuttals suppress collective action by driving wedges between groups within movements, a phenomenon we term the **wedging effect**. While government-led rebuttals may not necessarily convince everyone of their validity, they can deepen existing divides or create new ones, effectively fragmenting collective movements. This fragmentation undermines the solidarity essential for online activism. The wedging effect operates alongside a **chilling effect**, particularly among dissidents. Once the government sets the tone through an official rebuttal, dissidents may fear deviating from the shifting majority opinion or the government's stance. This fear creates normative pressure within activists, rendering it politically or socially unacceptable to publicly contradict the government. Consequently, this shift in the opinion dynamic creates an environment where dissent is muted, and online activism loses momentum.

We test this theory by analyzing an event of online activism in China as a natural experiment. Using a regression discontinuity design (RDD) coupled with computational analyses, we find that, surprisingly, government rumor rebuttal did not significantly alter people's political attitudes or factual judgments. However, these rebuttals have curtailed

information dissemination behaviors, especially among dissidents. The results further reveal that government rebuttals increased opinion divergence and emotional deviation among activists. This growing deviation between dissidents and the broader opinion climate ultimately suppressed their activism.

This study aims to contribute to the literature on misinformation, online activism, and digital repression in multiple aspects. First, our study enriches the theoretical discourse on rumor rebuttal by introducing the concepts of wedging and chilling effects, thus expanding beyond the conventional focus on credibility and persuasion. It challenges the prevailing notion that official rumor rebuttals primarily reshape factual perceptions or rectify public misbeliefs. Instead, we show that state-led narratives exert substantial influence by fragmenting opposition and dampening dissident engagement, even when people's underlying attitudes and factual understandings remain unchanged. In doing so, we expand the analytical lens beyond traditional persuasion paradigms and illuminate the subtle mechanisms through which state actors wield communicative power.

Second, this study advances the literature on online activism by highlighting a dual vulnerability in digital activism: while digital networks facilitate rapid mobilization, they are also susceptible to subtle state-driven information strategies that undermine collective resistance. Our findings show a state-driven "discursive engineering" that does not hinge on widespread acceptance of official rebuttal. Rather, the imposition of a particular narrative environment triggers "wedging" and "chilling" effects, effectively curtailing online mobilization. Thirdly, by combining machine learning models with a regression discontinuity design, this study provides robust evidence of these non-persuasion-based mechanisms. It aims

to contribute to the methodological toolbox that allows for the examination of digital activism in a real-world setting by exploiting naturally occurring variations in large-scale data. The methodology also facilitates the identification of causal relationships in a more dynamic and ecologically valid context.

2. Effects of Government-led Rumor Rebuttal on Online Activism

The question of whether, and if so how, government-led rumor rebuttals effectively mitigate online activism remains a critical area of inquiry. One prominent perspective suggests that rumor rebuttals from credible sources, backed by factual evidence, can correct misperceptions. For instance, studies on election rumors have shown that, under specific strategies, rumor rebuttal can effectively alter people's factual judgments, irrespective of their political attitudes (Aird et al., 2018; Clayton et al., 2020). Research has also confirmed that robust and vivid evidence can diminish public support for the content of the rumor (Huang, 2017). Yet, when the government debunks a rumor only to later contradict itself by realizing the rumor, public trust in the rebuttal decreases, along with overall satisfaction with the government (Wang & Huang, 2021).

Government rumor rebuttals can further transform political attitudes concerning contentious events through *persuasion*. Propaganda studies have shown that governments can bolster public understanding and garner support, thereby reducing the public's inclination to engage in protests (Adena et al., 2015; Cantoni et al., 2017; Gehlbach & Sonin, 2014; Guriev & Treisman, 2020; Peisakhin & Rozenas, 2018; Rozenas & Stukal, 2019). Furthermore, rumor rebuttal can shift the standards by which people evaluate controversial issue through the well-

documented *framing effect*. For instance, a study on the British Parliament found that when the "right person" presents the "right view," rumor rebuttals significantly increase the public's rejection rate of rumors, regardless of partisan preferences (Berinsky, 2017).

A contrasting view, however, suggests that activists may continue to discuss contentious issues, irrespective of the factual accuracy of the information governments provided. This view contends that online activism often stems from long-term grievances rather than being solely driven by misinformation, making it resistant to resolution through rumor rebuttals (Ahmed et al., 2017; Munger et al., 2019; Theocharis et al., 2015). Long-term factors such as economic crises, sustained inflation, and corruptions alongside short-term shocks like the sudden death or assaults on public figures can also provoke activism (Bennett & Segerberg, 2012; Munger et al., 2019; Pan & Siegel, 2020). Misinformation can thus serve as a deliberate strategy employed by opposition groups (Steinert-Threlkeld et al., 2015; Tufekci & Wilson, 2012). Even when specific misinformation is corrected, activists may advocate for broader changes.

Moreover, preexisting attitudes play a far more substantial role in shaping the interpretation of information provided by the government, often overshadowing the influence of factual corrections. People tend to interpret official rumor rebuttal through the lens of pre-existing beliefs, rather than allowing the rebuttals to reshape beliefs. This psychological perspective finds support in a large body of literature on *motivated reasoning,* which suggests that individuals tend to perceive information aligning with their judgments as factual while disregarding contradictory evidence (Barnidge et al., 2020; Bartels, 2002; Jerit & Barabas, 2012; Nir, 2011; Theodoridis, 2017). Political ideologies wield significant influence over individuals' interpretion of information (Crowder-Meyer & Ferrín, 2021; Webster &

Abramowitz, 2017). People are more inclined to endorse policies or information supported by their own party (Malka & Lelkes, 2010). These entrenched partisan attitudes can impede activists from accepting corrections to their misperceptions.

3. The Wedging and Chilling Effects of Rumor Rebuttal

Building upon the existing literature, we propose a theoretical framework arguing that government rebuttal – providing factual information – can suppress online activism without necessarily altering individuals' political attitudes or factual judgments. This suppression occurs through the creation of discrepancies among activists, where perceived opinion deviations from others discourage activism. Below, we elaborate on this theory.

The Wedging Effect: Fragmenting Solidarity

Our theory partially builds on the selective interpretation perspective. As discussed earlier, individuals interpret government rebuttals through the lens of their preexisting beliefs. When the government issues a rebuttal, those with moderate or pro-government views may find the rebuttal credible, prompting them to either adjust their opinions or disengage from activism. Conversely, critical citizens are more inclined to reject the rebuttal, viewing it as part of the government's broader effort to control the narrative. This selective acceptance of government rebuttals creates a wedge between different factions of activists, widening the gap between the pro-government activists, who primarily focus on the contentious event, and dissidents, whose activism may stem from deeper, long-standing grievances.

The wedging effect undermines online activism by fracturing movements from within. Rather than uniting around a common set of factual understandings, government rebuttals can

polarize opinions, creating internal divisions that erode the collective power of activist movements. This fragmentation undermines solidarity, as factions with differing interpretations struggle to maintain a shared factual perception. Without this unity, online activism becomes less effective in sustaining large-scale actions.

The Chilling Effect: Suppressing Dissent

The chilling effect alongside the wedging effect, deterring individuals from sharing information or expressing their opinions. Research has long substantiated this chilling effect, demonstrating that even threats unrelated to politics, such as fear triggered by images of danger, have been shown to reduce both hypothetical and actual protest behavior, along with the reduced expectation of others participating in protests (Young, 2019). Government propaganda often functions as a demonstration of state capacity, prompting individuals to conceal their political views to avoid potential consequences (Huang, 2015). Propaganda may not necessarily change views but can influence individuals' perception of others' willingness to protest, thereby reducing their tendency to protest (Huang & Cruz, 2022).

After the government issues a rebuttal, people refrains from activism, not because they are persuaded, but because they interpret the rebuttal as a signal of the government's stance and repressive capacity. The signal can lead individuals to self-censor, fearing social isolation or backlash for expressing dissenting opinions. This effect is particularly pronounced among the dissidents, who, after a government rebuttal establishes the dominant narrative, perceive a misalignment between their own views and the prevailing opinion climate. According to the "influence of presumed influence" theory, individuals presume that information has influenced others, and this presumption, in turn, affects their own behavior (Gunther & Storey, 2003). In

the context of government rebuttals, this perceived divergence from the opinion climate can foster the belief that others have accepted the rebuttal. This, in turn, diminishes the perceived willingness of others to participate in protest.

Moreover, government rebuttals can shape perceived social norms, compelling individuals—especially critical citizens—to avoid sharing information that contradicts the government's narrative. Research shows that government information release may also lead the public to mimic the expressions and behaviors of those who trust the government, creating an appearance of believing in the information released by the government (Little, 2017). This behavior reinforces the appearance of alignment with the official stance and discourages the spread of messages that challenge the official narrative. In this way, individuals may reduce their information sharing behaviors even if their factual beliefs and attitudes remain unchanged. This leads to our hypotheses:

Hypothesis 1: Government rumor rebuttal can significantly supress online activism.

Hypothesis 2: Perceived opinion discrepancies from others significantly reduces information dissemination behavior, especially among the dissidents.

Furthermore, numerous studies highlight the pivitol role of emotion, such as fear or anger, in stimulating online activism (Egorov et al., 2009; Gehlbach & Sonin, 2014; Mehdi & Dan, 2015). Following the same logic, government rumor rebuttal may elicit varying emotional responses and increase emotional divergence among the activists. For some, the clarification of truth may alleviate anger, while others may have to adjust emotional expressions in response to observed shifts in public sentiments (Barsade, 2002; Hareli & Rafaeli, 2008; Parkinson, 2011). Protesters often recalibrate their emotional responses by observing emotional cues from

their peers after a government rebuttal. They interpret emotional shifts within the activists and adjust their emotional expressions to align with the updated information presented in the rebuttal. Therefore, we put forward the final hypothesis:

Hypothesis 3: Government rumor rebuttal can lead to an increase in sentimental discrepancies among activists, which significantly reduces online activism.

4. Empirical Strategies

4.1 Research Design

We utilize an online collective protest as a natural experiment to test our hypotheses. Sina Weibo, often referred to as China's version of Twitter, was selected as it serves as a primary platform for online activism. Our research design exploits the discontinuity in the treatment variable – the release of a government-led rumor rebuttal – to identify its causal effect on the outcome of interest, namely online activism, using a regression discontinuity design (RDD). The online activism was triggered by a real-world event: a food safety incident at Chengdu No.7 Middle School (see appendix A1 for event profiles). This incident fueled intense online discussions but was later debunked following a local government investigation that provided evidence proving the claims to be false.

4.2 Empirical Specification

The online activism sparked by this incident meets a time discontinuity design(Hausman & Rapson, 2018), as depicted in Figure 1. The treatment assignment is determined by the timing of the government's rumor rebuttal. We treat netizens who posted after the government released the official rumor rebuttal (Phase 2) as the experimental group and those who posted before the

reversal point (Phase 1) as the control group. Time serves as an appropriate running variable in this event. On March 17, at 10:39 am, “Sichuan Online”, the social media account of Sichuan Daily, became the first to announce a news release from the Chengdu Joint Investigation Team online, which confirmed the rumor to be fake. Subsequently, at 11 am, numerous new media accounts on social media platforms began to report the rebuttal. Hence, this study designates 11:00 on March 17 as the cutoff point. The indicator variable of the RDD is defined as follows:

$$Di = \begin{cases} 1, & Zi \geq 0 \\ 0, & Zi < 0 \end{cases}, \quad (1)$$

where Zi represents the running variable, which denotes the number of days between a specific time and the critical point. For instance, if a person posts on Weibo about this event at 11:00am on March 14, then $Zi = -3$; if a person posts a related post at 11:00 on March 19, then $Zi = 2$. Consequently, $Zi > 0$ indicates that the person posted the Weibo after the government debunked the rumor, while $Zi \leq 0$ indicates that the person posted the Weibo before the government debunked the rumor. Di is the treatment status variable: $Di = 1$ for posts in the experimental group influenced by the rebuttal; and $Di = 0$ for those in the control group. Given that equation (1) holds, the RDD model writes as follows:

$$Yi = \alpha + \rho Di + f(zi, Di) + \gamma Xi + \varepsilon, \quad (2)$$

where $f(zi, Di)$ is the polynomial function or non-parametric form of the running variable Zi and the grouping variable Di, and Xi is the control variable. The coefficient ρ denotes the effect of government rumor rebuttal on protesters’ behavior at the breakpoint, which is the parameter this study aims to estimate. Please refer to Appendix A2 for a description of the pseudo-randomness assumption.

[Insert Figure 1 about here]

4.3 Data

We scraped all event-related posts and comments in real-time from March 13 to March 19, 2019, the entire duration of the online activism[①]. Using "Chengdu No.7 Middle School Food" and "No.7 Experimental School Food" as keywords, we collected the full sample of online discussions, yielding a total of 7,967 posts. Since many express their views on controversial events directly under the trending posts, we also collected all comments on the top 50 popular Weibo posts[②] using the same keywords, obtaining a total of 44,620 comments. In total, we collected 52,587 original posts and comments from 21,701 activists[③]. Following the initial steps of data cleaning, a total of 27,860 valid data entries were retained for analyses[④]. Figure 2 displays the daily changes in the volume of posts.

[Insert Figure 2 about here]

The natural experiment design necessitates the key assumption that the treatment group has indeed received the treatment. To validate this, we manually inspected posts from both the control group and the treatment group. First, the control group dataset was examined to ensure activists were not exposed to the rebuttal between 10:39 and 11:00 on March 17. For the experimental group, since the comments were derived from the top 50 popular Weibo posts—

① We conducted a second-round data collection to avoid potential spontaneous deletions and post-audit removals. The dataset spans from March 13, 2019, to February 26, 2020, and data that appeared repeatedly in the two collections was retained only once. Among them, data from March 13 to March 31, 2019, accounts for 98.3%, with only a very small number of new Weibo posts added afterward, indicating that this online collective action is essentially complete.

② The top 50 Weibo posts are ranked by popularity after search. These top 50 popular Weibo posts essentially cover all the highly discussed trending topics. The number of comments on the remaining Weibo posts does not exceed a dozen.

③ Each data point includes Weibo ID, content, number of comments, reposts, likes, pictures, original link, personal homepage, location, release time, user ID, username, gender, number of followers, fans, whether the user is a certified user, etc.

④ The data cleaning process involves removing words without real meaning, such as conjunctions and other stop words (using the stop word list from Ha'erbin Institute of Technology). Indecipherable content that cannot be decoded in a standard Chinese operating system was also removed. When the same user posts two or more identical Weibo posts simultaneously, only one was retained since the repetition may be attributed to network latency. Additionally, since Weibo has a "repost and comment" feature, and the majority of users choose to this feature when reposting, some reposts were collected twice. This study identified and recorded "repost and comment" based on the consistency of content, user, and time for each Weibo post, and removed all duplicates.

and all popular posts made after 11:00 contained the rumor rebuttal content— these comment data were considered to have been exposed to the rebuttal. A similar approach was applied to original Weibo posts, including those that forwarded the rumor rebuttal. Posts for which exposure could not be confirmed were excluded from the analysis. After this step, 26,948 posts were retained, accounting for 96.73% of the original dataset. This procedure allows us to construct a control group, consisting of activists who were unaware of the government rebuttal, and an experimental group, comprising those who were aware of it. This setup meets with the conditions for applying RDD, where treatment assignment is determined by a threshold— in this case, the timing of the rumor rebuttal.

4.4 Measurements

Next, we transform unstructured, low-value-density text data into structured, high-value-density analyzable variables using machine learning algorithms. Since a post may contain various information related to our variables of interest, we employ both supervised learning methods and unsupervised learning methods, as per the requirements of different variables. See Appendix Table B1-B4 for all detailed procedures and examples.

Information dissemination

Active engagement is central to online activism as it facilitates mobilization and solidarity-building. To measure the key outcome variable – information dissemination – we use several engagement metrics. These include reposting, mentioning relevant entities(@), tagging(#)[①] and explicit calls for attention, such as phrases such as “let’s all support” and “let’s all follow”. This

[①] Tagging serves as a crucial tool for mobilizing collective action. When a topic is tagged, it consolidates related content, facilitating the formation of a collective force in public opinion and amplifying the visibility of related content. Additionally, clicking or searching for the tag redirects users to the Weibo search results page, where they can directly access related posts. This mechanism non-followers to browse event-related posts, effectively achieving the goal of information dissemination.

variable evaluates whether users engage in these behaviors in their posts. If so, the subvariable value assigned is 1; if not, the subvariable value assigned is 0. *Information dissemination* variable is the sum of the values of four subvariables. We employ supervised machine learning to capture this variable, and the accuracy rate for the testing set was 94.00%.

Political Attitude

On the Chinese Internet, political attitudes range from nationalist (left) to liberal (right). Nationalists typically support the country and government, while liberals often express skepticism (Han, 2015). Given the diversity and subtlety in the expression of political attitudes, supervised learning models may struggle to effectively capture the full spectrum of political attitudes with a limited training dataset. Therefore, we employ unsupervised learning methods to extract political attitudes, reducing the potential influence of human biases. The term frequency-inverse document frequency (TF-IDF) algorithm is utilized to generate the text vector matrix representing each post. Subsequently, the K-means algorithm of unsupervised machine learning is utilized to cluster all text data into seven categories. Through a thorough analysis of these seven categories, we classify them into three overarching political attitudes: dissidents, neutral or unclear, and pro-government. These attitudes are then assigned numerical values of 1, 0, and -1, respectively[①]. This systematic approach enables us to effectively categorize and analyze netizens' political attitudes. We also employ the GPT-4 model for supervision and validation, achieving an accuracy rate of 78.23%.

Factual Perception

Activists' factual perceptions were assessed by analyzing the content of their posts, i.e. whether

[①] To ensure the classification validity of unsupervised learning, a validity test is provided in the Appendix B.

they believed in the government rebuttal and which version of the truth they supported. This variable was extracted using a supervised learning algorithm, achieving an accuracy rate of 87.20% on the testing set. Following the classification spectrum outlined above, a value of 1 is assigned if the person concurs with the government's actions, 0 if no clear inclination or judgment is expressed, and -1 if the person criticizes the government and the public institutions.

Emotion

Sentiment analysis, performed via machine learning methods, allows for the extraction of emotional tone from the subjective text posted by the user. This study employs the BERT (Bidirectional Encoder Representations from Transformers) model, pre-trained on Chinese corpora. Leveraging extensive pre-training in Chinese, it extracts features from the text in context, considering both preceding and following text, to discern its contextual meaning. A random sample of 1500 data points was used to test the accuracy of the emotion label, yielding an accuracy rate of 82.27%[①].

Opinion Deviation

Opinion deviation gauges the extent to which a user's discourse deviates from the average online public opinion. It becomes relevant when people's views align or conflict with mainstream perspectives, leading to adjustments in their behaviors. We employ unsupervised machine learning to extract this variable. Similarly, the TF-IDF algorithm is utilized to generate the text vector matrix of each post. Subsequently, the text parameters are obtained by reducing the dimensionality via the PCA algorithm. These text parameters are then standardized to derive the degree of deviation of each post from the mean, resulting in the degree of opinion deviation.

[①] Netizens in online collective actions often use sarcasm for negative expression, which is difficult for machine learning models to recognize, and the accuracy rate is relatively low, as expected.

Finally, the opinion deviation degree is normalized.

$$\text{Standardization：}\ y = \frac{X_i - \mu}{\sigma}$$

$$\text{Normalization：}\ y = \frac{X_i - X_{min}}{X_{max} - X_{min}}$$

Emotion Deviation

We devised an "emotion deviation" variable that quantifies the extent to which an individual's emotion diverges from the overall emotional climate. Employing the locally weighted regression (LOWESS) smoothing technique, we performed a temporal emotion analysis to ascertain the average emotion at each temporal juncture, representing the collective emotional climate. We then calculated the disparity between each person's emotion and the collective emotional climate, standardizing, taking the absolute value, and normalizing these disparities sequentially. Table 1 provides descriptive statistics for the key variables extracted from textual data through data mining. See Appendix Table B5 for a complete list of descriptive statistics.

[Insert Table 1 about here]

5. Results

5.1 Government Rebuttal Fails to Change Factual Beliefs or Political Attitudes

We employ the Regression Discontinuity Design to analyze whether government rumor rebuttals lead to significant changes in engagement behaviors, political attitudes, factual perceptiona, and emotions. In line with the guidelines set by Lee and Lemieux (2010), we sequentially present graphics, non-parametric estimation results, parameter estimation results, and validity tests of our analyses. The optimal bandwidth h_0 for each model is calculated according to the mserd-method of Calonico et al. (2014).

[Insert Figure 3 about here]

Figure 3 illustrates the non-linear relationship between the running variable and the outcome variable on both sides of the cutoff point. While political attitude and factual judgment do not show clear discontinuities, emotion does. Tables 2 further present the non-parametric estimation results of the government's rumor rebuttal on these dependent variables, respectively. The results confirm that the government-led rumor rebuttal does not significantly affect factual judgment or political attitude[①]. The estimation for emotion is insignificant as well, indicating that rumor rebuttal does not alter average online sentiments either. Moreover, the signs and levels of significance of the non-parametric estimation coefficients align with those of the parameter estimation, suggesting that the estimations are robust and independent of the model's parameter settings[②].

[Insert Table 2 about here]

Balance Tests

One potential concern is the selective entry and exit of participants in online activism. If two different groups of individuals were active before and after the release of the rebuttal, it could raise concerns about the validity of our experimental framework. To address this, we conducted balance tests, the results of which are presented in Table 3. All covariates passed the balance test. The all above tests indicating that rebuttal has not caused users of specific attitudes, judgments, emotions or characteristics to disproportionately enter or exit the discussion. The

[①] Since not expressing a factual opinion is coded as 0, a null result could also reflect less expression of factual opinions, while the differences in proportions between 1 and -1 could actually be significantly changing. We tested this as a robustness test by deleting the samples where the factual judgment=0 and re-estimating. The result is again insignificant at the 10% level, indicating robustness.

[②] Due to word limitation, see Appendix Table C1-C7 for parameter estimation results. See Appendix D for results from a series of validity tests.

balance test serves a validity check to confirm that the RDD holds.

[Insert Table 3 about here]

5.2 Government Rebuttal Suppressed Online Activism

Contrarily, information dissemination behavior exhibits a distinct discontinuity at the breakpoint of the government's rumor rebuttal. Figure 4 illustrates the non-linear relationship between the running variable and the outcome variable on both sides of the cutoff point, providing preliminary evidence that the rumor rebuttal may have reduced online activism.

[Insert Figure 4 about here]

Tables 4 further presents the non-parametric estimation results of the government's rumor rebuttal on dissemination behavior. The results shows that government rumor rebuttal significantly reduced the dissemination behavior by 0.062, almost 50% at that moment. It is thus essential to further examine the regression discontinuity effect on dissemination behavior across different bandwidths.

[Insert Table 4 about here]

In Table 5, we report results from the non-parametric estimation under five bandwidths. The regression discontinuity effects remain significant at both the 5% and 1% levels across all bandwidths. This consistency suggests that the results are independent of the model's parameter settings and remain robust. The finding confirm that the government-led rumor rebuttal significantly reduced dissemination behavior, thereby lending strong support for Hypothesis 1.

[Insert Table 5 about here]

5.3 Wedging and Chilling

In this section, we examine how government rumor rebuttal suppresses online activism.

As shown in Table 5, heterogeneous analyses from both estimation models consistently show that reduction in information dissemination is significant among dissidents. This indicates that government rumor rebuttal can significantly reduce the dissemination behavior of the critical netizens, demonstrating the *chilling* effect in action. The effect of government rumor rebuttal on those with "pro-government" and "neutral or unclear" political attitudes is insignificant, contrarily.

Hypothesis 2 posits that individuals gauge the impact of rumor countering based on the perceived public opinion shift, which, in turn, influences their propensity to disseminate information. To test this, it is vital to examine whether the government rumor rebuttal siginificantly altered the average opinion deviation. As shown in Figure 5 and Table 6 reveal that government rumor rebuttal did not result in a universal shift in discourse. Instead, it significantly increased the deviation between dissidents and others, reinforcing a *wedging* effect.

[Insert Figure 5 about here]

[Insert Table 6 about here]

Subsequently, we employ a fuzzy-RDD with discourse deviation being the treatment variable and dissemination behavior as the dependent variable, to determine if the observed change in dissemination behavior among the dissidents can be attributed to their perception of altered opinion climate. As shown in Table 7, we find that opinion deviation accounts for the significant reduction in dissemination behavior among dissidents. By contrast, this effect is insignificant among those who hold pro-government or neutral attitudes. The results lend support for Hypothesis 2 (see Appendix Table C6 for complete heterogeneity results).

[Insert Table 7 about here]

Figure 6 further illustrates effect of rumor rebuttal on emotional deviation, suggesting that the rebuttal has led to increased disparities within online sentiments as well. As demonstrated in Table 8, we observe that rumor rebuttal significantly heightened emotion deviation, particularly among individuals with pro-government and neutral attitudes. However, this effect was not evident among the dissents. This finding indicates that while government rumor rebuttal failed to influence the emotional alignment within the dissidents, it successfully created sentimental shifts within the pro-governments and those with neutral political attitudes. It is important to note that the predominance of critical activists contributes to an overall negative emotional climate. However, the emotional solidarity among pro-government individuals and those with neutral attitudes is curtailed, leading them to adjust emotions to deviate from the predominantly negative sentiments. Therefore, Hypothesis 3 is partially supported.

[Insert Figure 6 about here]

Taken together, the results show that government rebuttals did not alter people's factual judgments or attitudes at an aggregative level. However, they did notably reduce information dissemination behavior, especially among critical citizens. This effect is driven by a combination of wedging and chilling effects: the increased opinion deviation between dissidents and others reduces activism, while government rebuttals encourage pro-government and neutral individuals to adopt emotions that diverge from the prevailing negative emotional climate, further distancing themselves emotionally from the critical citizens. These findings suggest that the cessation of online protests stems from activists' knowledge that the event has been debunked as a rumor, leading to a widening discrepancy between activists. This

diminished belief in others' willingness to protest together, combined with the chilling effect reduce people's inclination to continue protesting.

[Insert Table 8 about here]

6. Discussion

This study contributes to a long-standing debate on the efficacy of government rebuttals as a tactic for curbing online activism. Departing from traditional fact-driven and motivated reasoning models, we present a theoretical framework illustrating how government-led rumor rebuttals can suppress activism—not by altering core beliefs or factual perceptions, but by fragmenting opposition groups and cultivating a climate that discourages dissent. Our findings reveal that this fragmentation, coupled with a chilling effect, systematically erodes the solidarity and resolve of activist communities. As a result, collective resistance weakens, as internal divisions and a sense of futility undermine the willingness of dissidents to sustain their efforts. This perspective highlights a subtle yet powerful mechanism of digital repression that operates through social and psychological channels rather than outright coercion.

Our study broadens the discourse on digital activism by introducing a non-coercive, state-driven mechanism for reshaping the collective dynamics of online movements. Operating through the dual forces of "wedging" and "chilling," government-led rumor rebuttals do not persuade through the force of reasoned argumentation, nor do they rely on overt repression. Instead, they strategically exploit group vulnerabilities, fostering internal divisions and discouraging individual participation. By eroding solidarity within activist networks, these rebuttals make collective actions harder to sustain and ultimately diminish the influence of the movement. In this sense, government interventions not only rearrange the terrain of digital

activism but also recalibrate the emotional and cognitive underpinnings of dissent, with profound implications for the trajectory and efficacy of collective resistance.

Our findings also extend the literature on misinformation by unveiling the distinct mechanisms through which government-led rumor rebuttals operate. Rather than primarily aiming to correct misperceptions or shift underlying attitudes, these interventions strategically target the social fabric of opposition networks. By signaling the state's capacity and suggesting that others have embraced the official narrative, rebuttals breed uncertainty among would-be activists. As a result, potential dissidents hesitate to sustain their efforts, anticipating diminished collective support and heightened personal risk. This process of discursive engineering—dividing opposition groups and discouraging active participation—effectively stifles online activism without the need for overt coercion. In this sense, government rebuttals function less as tools of persuasion and more as subtle instruments of demobilization, reshaping the digital political arena by eroding solidarity and suppressing collective action.

This research contributes to the methodological toolbox for studying digital politics by combining causal inference and computational analyses. The study employs machine learning methods for data mining, obtaining variables such as political attitudes and opinion deviation, which can be challenging to obtain by traditional methods. This provides useful references for data mining in information politics and political behavior research. The study also bears several limitations: although we are able to detect changes in the opinion climate, the nature of social media's free entry and exit dynamics limits our ability to draw individual-level inferences within this natural experiment setting. We use balance tests to minimize this limitation as effectively as possible. Second, like many social media studies, this research faces the potential

issue of censorship. However, much of our analysis relies on comments and posts directly under the original rebuttal posts, which partially mitigates this concern to some extent. Third, social media posts often contain nuanced expressions such as irony and sarcasm. Given the large volumne of data, it is challenging to identify and eliminate potential errors. Future studies may benefit from more sophisticated natural language processing techniques or manual annotation to better capture such nuances and enhance the robustness of the analysis.

While social media platforms are favored for their lower costs and ability to organize collective actions (Zhuravskaya et al., 2020), our study highlights their susceptibility to government information maneuvers. In response to government rebuttals, speculation about others' engagement may diminish one's inclination to act. Regardless of whether people believe in the factual accuracy of the rebuttal, government rebuttal can be effective in causing activists to believe that rebuttal affects the protest intentions of fellow netizens, thereby reducing their own behaviors. Activists may also reduce their participation in online discussions or protests—not necessarily because they are persuaded by the government's correction, but because they become aware of the line drawn by the government. The chilling effect significantly impact online activism, particularly as information-sharing behaviors like retweeting are critical for stimulating participation (Boyd et al., 2010). A substantial reduction in visibal engagement diminishes the attention resources necessary for maintaining collective action, ultimately leading to its gradual decline (Hunt & Gruszczynski, 2021; Tufekci, 2013). The "social" nature of online platforms can be both advantageous and detrimental.

Data Availability

The authors agree to make all data and script files publicly available.

AI Statement

In accordance with the Taylor & Francis AI Policy, we confirm that we have used ChatGPT 4, during the preparation of this manuscript. Specifically, we employed this tool for proofreading. We have disclosed this usage in the manuscript and ensured that it complies with the guidelines set out by the Taylor & Francis AI Policy.

Reference

Adena, M., Enikolopov, R., Petrova, M., Santarosa, V., & Zhuravskaya, E. (2015). Radio and the rise of the nazis in prewar Germany. *The Quarterly Journal of Economics*, *130*(4), 1885-1939. https://doi.org/10.1093/qje/qjv030

Ahmed, S., Jaidka, K., & Cho, J. (2017). Tweeting India's Nirbhaya protest: A study of emotional dynamics in an online social movement. *Social Movement Studies*, *16*(4), 447-465. https://doi.org/10.1080/14742837.2016.1192457

Aird, M. J., Ecker, U. K. H., Swire, B., Berinsky, A. J., & Lewandowsky, S. (2018). Does truth matter to voters? The effects of correcting political misinformation in an Australian sample. *Royal Society Open Science*, *5*(12), 180593. https://doi.org/doi:10.1098/rsos.180593

Barnidge, M., Gunther, A. C., Kim, J., Hong, Y., Perryman, M., Tay, S. K., & Knisely, S. (2020). Politically motivated selective exposure and perceived media bias.

Communication Research, *47*(1), 82-103. https://doi.org/10.1177/0093650217713066

Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly*, *47*(4), 644-675. https://doi.org/10.2307/3094912

Bartels, L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior*, *24*(2), 117-150. https://doi.org/10.1023/A:1021226224601

Bennett, W. L., & Segerberg, A. (2012). The logic of connective action: Digital media and the personalization of contentious politics. *Information, Communication & Society*, *15*(5), 739-768. https://doi.org/10.1080/1369118X.2012.670661

Berinsky, A. J. (2017). Rumors and health care reform: Experiments in political misinformation. *British Journal of Political Science*, *47*(2), 241-262. https://doi.org/10.1017/S0007123415000186

Boyd, D., Golder, S., & Lotan, G. (2010). *Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter* 2010 43rd Hawaii International Conference on System Sciences (HICSS-43), Honolulu, HI, USA.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, *82*(6), 2295-2326. https://doi.org/10.3982/ECTA11757

Cantoni, D., Chen, Y., Yang, D. Y., Yuchtman, N., & Zhang, Y. J. (2017). Curriculum and ideology. *Journal of Political Economy*, *125*(2), 338-392. https://doi.org/10.1086/690951

Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., Kawata, A., Kovvuri,

A., Martin, J., Morgan, E., Sandhu, M., Sang, R., Scholz-Bright, R., Welch, A. T., Wolff, A. G., Zhou, A., & Nyhan, B. (2020). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior*, *42*(4), 1073-1095. https://doi.org/10.1007/s11109-019-09533-0

Crowder-Meyer, M., & Ferrín, M. (2021). The effects of ideological and ethnoracial identity on political (mis) information. *Public Opinion Quarterly*, *85*(3), 753-779. https://doi.org/10.1093/poq/nfab038

Earl, J., Maher, T. V., & Pan, J. (2022). The digital repression of social movements, protest, and activism: A synthetic review. *Science Advances*, 8(10), eabl8198. https://doi.org/doi:10.1126/sciadv.abl8198

Egorov, G., Guriev, S., & Sonin, K. (2009). Why resource-poor dictators allow freer media: A theory and evidence from panel data. *American Political Science Review*, 103(4), 645-668. https://doi.org/10.1017/S0003055409990219

Gehlbach, S., & Sonin, K. (2014). Government control of the media. *Journal of Public Economics*, *118*, 163-171. https://doi.org/10.1016/j.jpubeco.2014.06.004

Gunther, A. C., & Storey, J. D. (2003). The influence of presumed influence. *Journal of Communication*, *53*(2), 199-215. https://doi.org/10.1093/joc/53.2.199

Guriev, S., & Treisman, D. (2020). A theory of informational autocracy. *Journal of Public Economics*, *186*, 104158. https://doi.org/10.1016/j.jpubeco.2020.104158

Han, R. (2015). Defending the authoritarian regime online: China's “voluntary fifty-cent army”. *The China Quarterly*, *224*, 1006-1025.

https://doi.org/10.1017/S0305741015001216

Hareli, S., & Rafaeli, A. (2008). Emotion cycles: On the social influence of emotion in organizations. *Research in Organizational Behavior*, *28*, 35-59. https://doi.org/10.1016/j.riob.2008.04.007

Hausman, C., & Rapson, D. S. (2018). Regression discontinuity in time: Considerations for empirical applications. *Annual Review of Resource Economics*, *10*, 533-552. https://doi.org/10.1146/annurev-resource-121517-033306

Hobbs, W. R., & Roberts, M. E. (2018). How sudden censorship can increase access to information. *American Political Science Review*, *112*(3), 621-636. https://doi.org/10.1017/S0003055418000084

Huang, H. (2015). Propaganda as signaling. *Comparative Politics*, *47*(4), 419-444. https://doi.org/10.5129/001041515816103220

Huang, H. (2017). A war of (mis) information: The political effects of rumors and rumor rebuttals in an authoritarian country. *British Journal of Political Science*, *47*(2), 283-311. https://doi.org/10.1017/S0007123415000253

Huang, H., & Cruz, N. (2022). Propaganda, presumed influence, and collective protest. *Political Behavior*, *44*(4), 1789-1812. https://doi.org/10.1007/s11109-021-09683-0

Hunt, K., & Gruszczynski, M. (2021). The influence of new and traditional media coverage on public attention to social movements: The case of the Dakota Access Pipeline protests. *Information, Communication & Society*, *24*(7), 1024-1040. https://doi.org/10.1080/1369118X.2019.1670228

Jansen, S. C., & Martin, B. (2015). The streisand effect and censorship backfire. *International*

Journal of Communication, *9*, 656-671.

Jerit, J., & Barabas, J. (2012). Partisan perceptual bias and the information environment. *The Journal of Politics*, *74*(3), 672-684. https://doi.org/10.1017/S0022381612000187

King, G., Pan, J., & Roberts, M. E. (2017). How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. *American Political Science Review*, *111*(3), 484-501. https://doi.org/10.1017/S0003055417000144

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*(2), 281-355. https://doi.org/10.1257/jel.48.2.281

Little, A. T. (2017). Propaganda and credulity. *Games and Economic Behavior*, *102*, 224-232. https://doi.org/10.1016/j.geb.2016.12.006

Malka, A., & Lelkes, Y. (2010). More than ideology: Conservative–liberal identity and receptivity to political cues. *Social Justice Research*, *23*(2), 156-188. https://doi.org/10.1007/s11211-010-0114-3

Mehdi, S., & Dan, B. (2015). State censorship. *American Economic Journal: Microeconomics*, *7*(2), 280-307. https://doi.org/10.1257/mic.20130221

Munger, K., Bonneau, R., Nagler, J., & Tucker, J. A. (2019). Elites tweet to get feet off the streets: Measuring regime social media strategies during protest. *Political Science Research and Methods*, *7*(4), 815-834. https://doi.org/10.1017/psrm.2018.3

Nir, L. (2011). Motivated reasoning and public opinion perception. *Public Opinion Quarterly*, *75*(3), 504-532. https://doi.org/10.1093/poq/nfq076

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303-330. https://doi.org/10.1007/s11109-

010-9112-2

Pan, J., & Siegel, A. A. (2020). How Saudi crackdowns fail to silence online dissent. *American Political Science Review*, *114*(1), 109-125. https://doi.org/10.1017/S0003055419000650

Parkinson, B. (2011). Interpersonal emotion transfer: Contagion and social appraisal. *Social and Personality Psychology Compass*, *5*(7), 428-439. https://doi.org/10.1111/j.1751-9004.2011.00365.x

Peisakhin, L., & Rozenas, A. (2018). Electoral effects of biased media: Russian television in Ukraine. *American Journal of Political Science*, *62*(3), 535-550. https://doi.org/10.1111/ajps.12355

Roberts, M. (2018). *Censored: Distraction and diversion inside China's Great Firewall*. Princeton University Press. https://doi.org/10.1017/S1537592718002608

Rozenas, A., & Stukal, D. (2019). How autocrats manipulate economic news: Evidence from Russia's state-controlled television. *The Journal of Politics*, *81*(3), 982-996. https://doi.org/10.1086/703208

Steinert-Threlkeld, Z. C., Mocanu, D., Vespignani, A., & Fowler, J. (2015). Online social networks and offline protest. *EPJ Data Science*, *4*(1), 19. https://doi.org/10.1140/epjds/s13688-015-0056-y

Stukal, D., Sanovich, S., Bonneau, R., & Tucker, J. A. (2017). Detecting bots on Russian political Twitter. *Big Data*, *5*(4), 310-324. https://doi.org/10.1089/big.2017.0038

Theocharis, Y., Lowe, W., van Deth, J. W., & García-Albacete, G. (2015). Using Twitter to mobilize protest action: Online mobilization patterns and action repertoires in the

Occupy Wall Street, Indignados, and Aganaktismenoi movements. *Information, Communication & Society*, *18*(2), 202-220. https://doi.org/10.1080/1369118X.2014.948035

Theodoridis, A. G. (2017). Me, myself, and (I),(D), or (R)? Partisanship and political cognition through the lens of implicit identity. *The Journal of Politics*, *79*(4), 1253-1267. https://doi.org/10.1086/692738

Tufekci, Z. (2013). "Not this one": Social movements, the attention economy, and microcelebrity networked activism. *American Behavioral Scientist*, *57*(7), 848-870. https://doi.org/10.1177/0002764213479369

Tufekci, Z., & Wilson, C. (2012). Social media and the decision to participate in political protest: Observations from tahrir square. *Journal of Communication*, *62*(2), 363-379. https://doi.org/10.1111/j.1460-2466.2012.01629.x

Wang, C., & Huang, H. (2021). When "fake news" becomes real: The consequences of false government denials in an authoritarian country. *Comparative Political Studies*, *54*(5), 753-778. https://doi.org/10.1177/0010414020957672

Webster, S. W., & Abramowitz, A. I. (2017). The ideological foundations of affective polarization in the US electorate. *American Politics Research*, *45*(4), 621-647. https://doi.org/10.1177/1532673X17703132

Woolley, S. C., & Howard, P. N. (2016). Political communication, computational propaganda, and autonomous agents. *International Journal of Communication*, *10*, 4882–4890.

Yang, G. (2009a). China since Tiananmen: online activism. *Journal of Democracy*, 20(3), 33-36. https://doi.org/10.1353/jod.0.0094

Yang, G. (2009b). *The Power of the Internet in China: Citizen Activism Online*. Columbia University Press.

Yang, G. (2018). Demobilizing the Emotions of Online Activism in China: A Civilizing Process. *International Journal of Communication*, 12, 1945-1965.

Young, L. E. (2019). The psychology of state repression: Fear and dissent decisions in Zimbabwe. *American Political Science Review*, *113*(1), 140-155. https://doi.org/10.1017/S000305541800076X

Zhuravskaya, E., Petrova, M., & Enikolopov, R. (2020). Political effects of the internet and social media. *Annual Review of Economics*, *12*, 415-438. https://doi.org/10.1146/annurev-economics-081919-050239

Tables

Table 1 Descriptive statistics

Key variables	N	Mean	Standard error	Minimum	Maximum
Political Attitude	26,948	0.573	0.778	-1	1
Factual Perception	26,948	-0.457	0.646	-1	1
Emotion	26,948	0.316	0.471	-1	1
Dissemination Behavior	26,948	0.191	0.516	0	4
Opinion Deviation	26,948	0.101	0.126	0	1
Emotion Deviation	26,948	0.159	0.153	0.019	1

Table 2 Non-parametric Estimation of Government Rumor Rebuttal

	Political Attitude	Factual Judgment	Emotion
	0.042	0.024	0.041
	(0.054)	(0.035)	(0.025)
Optimal bandwidth	0.688	1.308	0.796
Effective N	7302	10369	7684
N	26948	26948	26948
Covariates controlled	-0.023	0.025	0.040
	(0.044)	(0.035)	(0.025)
Optimal bandwidth	0.956	1.359	0.780
Effective N	8775	10470	7628
N	26948	26948	26948

*p < 0.1, **p < 0.05, ***p < 0.01, standard errors in parentheses. *Note* The kernel function used for local polynomial inference is the triangular kernel density function.

Table 3 Balance Tests

	(1) Influence	(2) Certified Account	(3) Official Account
	-114,766.579	0.011	-0.002
	(393,586.544)	(0.021)	(0.005)
Optimal bandwidth	1.462	0.434	0.359
Effective N	10828	6070	5561
N	26948	26948	26948

*p < 0.1, **p < 0.05, ***p < 0.01, standard errors in parentheses. *Note* The kernel function used for local polynomial inference is the triangular kernel density function.

Table 4 Non-parametric Estimation of Government Rumor rebuttal on Dissemination Behavior

	Dissemination Behavior
	-0.062**
	(0.026)
Optimal bandwidth	1.209
Effective N	10106
N	26948
Covariates controlled	-0.062***
	(0.020)
Optimal bandwidth	1.667
Effective N	12662
N	26948

*p < 0.1, **p < 0.05, ***p < 0.01, standard errors in parentheses. *Note* The kernel function used for local polynomial inference is the triangular kernel density function.

Table 5 Non-parametric Estimation of Government Rumor Rebuttal on Dissemination Behavior

	h_0=1.667	h_1=0.8h_0	h_2=1.2h_0	h_3=1.5h_0	h_4=2h_0
All	-0.062***	-0.062***	-0.058***	-0.053***	-0.024**
	(0.020)	(0.024)	(0.016)	(0.015)	(0.012)
Effective N	12662	10414	17686	19261	24050
N	26948	26948	26948	26948	26948
Dissidents	-0.052***	-0.050***	-0.057***	-0.051***	-0.026**
	(0.017)	(0.019)	(0.014)	(0.013)	(0.011)
Effective N	9338	7572	13231	14524	18277
N	26948	26948	26948	26948	26948
Other Political Attitudes	-0.094	-0.092	-0.067	-0.050	0.064
	(0.066)	(0.076)	(0.054)	(0.051)	(0.041)
Effective N	3324	2842	4455	4737	5773
N	26948	26948	26948	26948	26948

*p < 0.1, **p < 0.05, ***p < 0.01, standard errors in parentheses. *Note* The kernel function used for local polynomial inference is the triangular kernel density function. Covariates controlled. Optimal bandwidth is h_0.

Table 6 Non-parametric Estimation of Government Rumor rebuttal on Discourse Deviation

	h_0=1.557	0.8h_0	1.2h_0	1.5h_0	2h_0
All	-0.001	0.002	-0.003	-0.001	0.004
	(0.006)	(0.006)	(0.005)	(0.004)	(0.004)
Effective N	11510	10189	16827	18702	23865
N	26948	26948	26948	26948	26948
Dissidents	0.012**	0.018***	0.009*	0.010**	0.012***
	(0.005)	(0.006)	(0.005)	(0.004)	(0.004)
Effective N	5286	4860	5625	6387	7410
N	26948	26948	26948	26948	26948

*p < 0.1, **p < 0.05, ***p < 0.01, standard errors in parentheses. *Note* The kernel function used for local polynomial inference is the triangular kernel density function. Covariates controlled. Optimal bandwidth is h_0.

Table 7 Non-parametric Estimation of Content Deviation on Dissemination Behavior

	h_0=1.072	0.8h_0	1.2h_0	1.5h_0	2h_0
Dissidents	-2.494**	-1.988**	-3.047*	-4.430*	-5.773*
	(1.187)	(0.906)	(1.584)	(2.578)	(3.132)
Effective N	6972	5982	7052	8755	13924
N	26948	26948	26948	26948	26948

*p < 0.1, **p < 0.05, ***p < 0.01, standard errors in parentheses. *Note* The kernel function used for local polynomial inference is the triangular kernel density function. Covariates controlled. Optimal bandwidth is h_0.

Table 8 Non-parametric Estimation of Government Rumor Rebuttal on Emotion Deviation

	h_0=0.940	0.8h_0	1.2h_0	1.5h_0	2h_0
All	0.018**	0.016*	0.024***	0.027***	0.032***
	(0.009)	(0.010)	(0.008)	(0.008)	(0.006)
Effective N	8622	7547	9844	10630	16864
N	26948	26948	26948	26948	26948
Dissidents	0.010	0.008	0.014	0.019**	0.027***
	(0.011)	(0.011)	(0.010)	(0.009)	(0.007)
Effective N	6406	5637	7189	7713	12531
N	26948	26948	26948	26948	26948
Other Political Attitudes	0.042**	0.039**	0.046***	0.047***	0.047***
	(0.017)	(0.018)	(0.016)	(0.016)	(0.012)
Effective N	2216	1910	2655	2917	4333
N	26948	26948	26948	26948	26948

*p < 0.1, **p < 0.05, ***p < 0.01, standard errors in parentheses. *Note* The kernel function used for local polynomial inference is the triangular kernel density function. Covariates controlled. Optimal bandwidth is h_0.

Figures

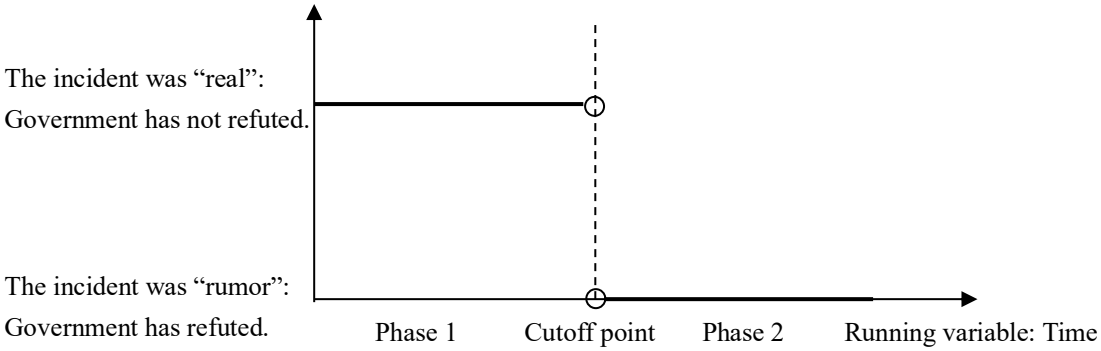

Figure 1 Diagram of Discontinuity Design

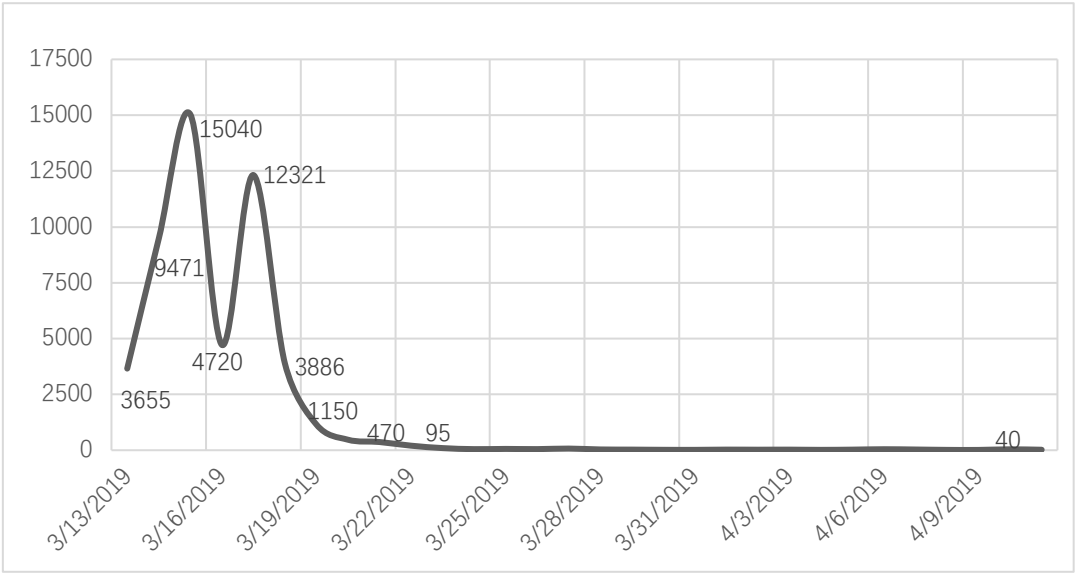

Figure 2 Volume of Daily Posts on the Chengdu No.7 Middle School Food Safety Incident

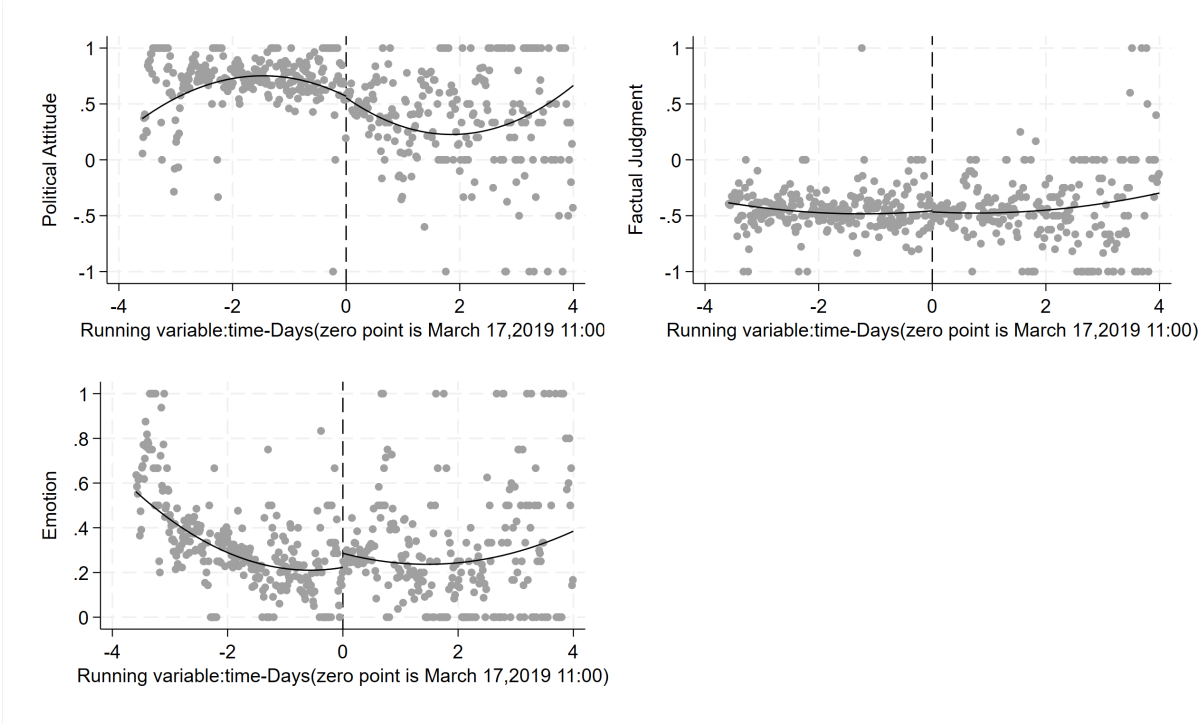

Figure 3 The Impact of Government Rumor Rebuttal on Activists' Political Attitude, Factual Judgement and Emotion

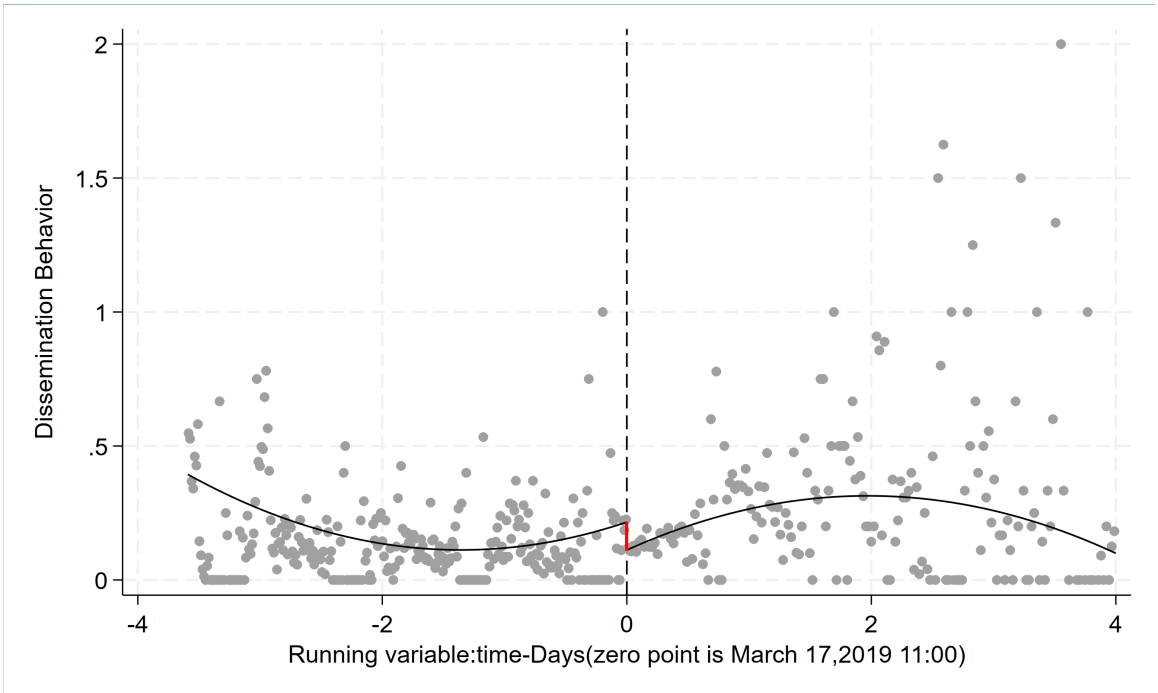

Figure 4 The Impact of Government Rumor Rebuttal on Activists' Dissemination Behavior

Note The fitting line represents a quadratic fit of the dissemination behavior to the running variable, with covariates controlled.

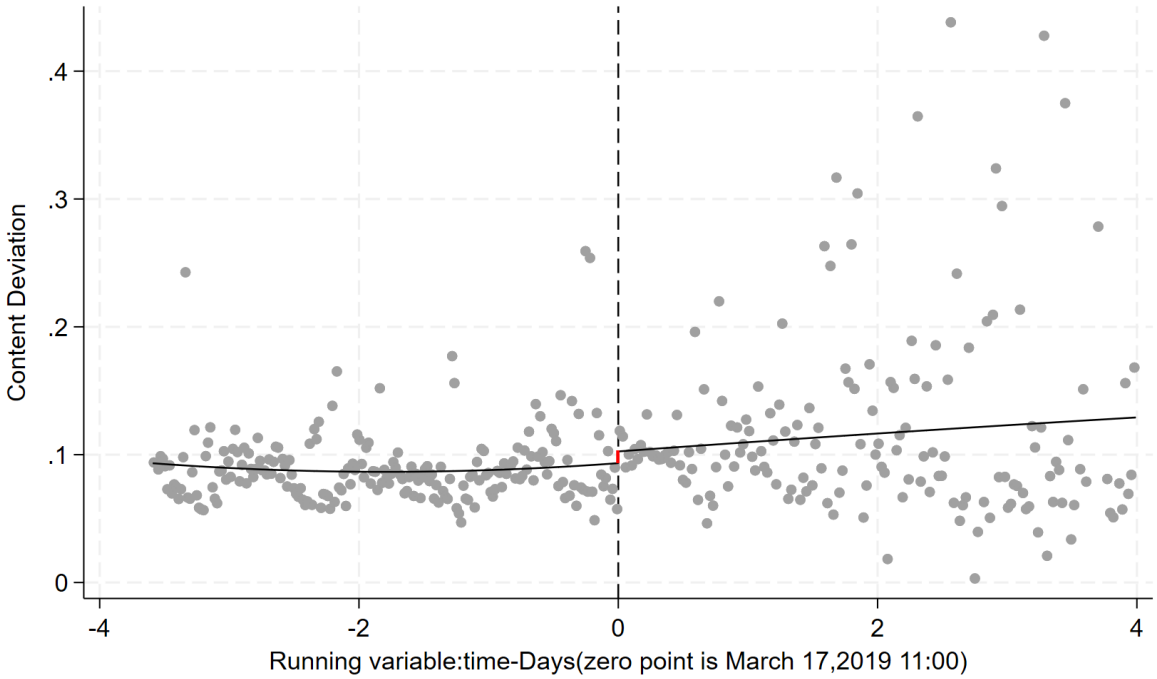

Figure 5 The Impact of Government Rumor Rebuttal on the Dissidents' Discourse Deviation

Note The fitting line represents a quadratic fit of the Dissidents' discourse deviation to the running variable, with covariates controlled.

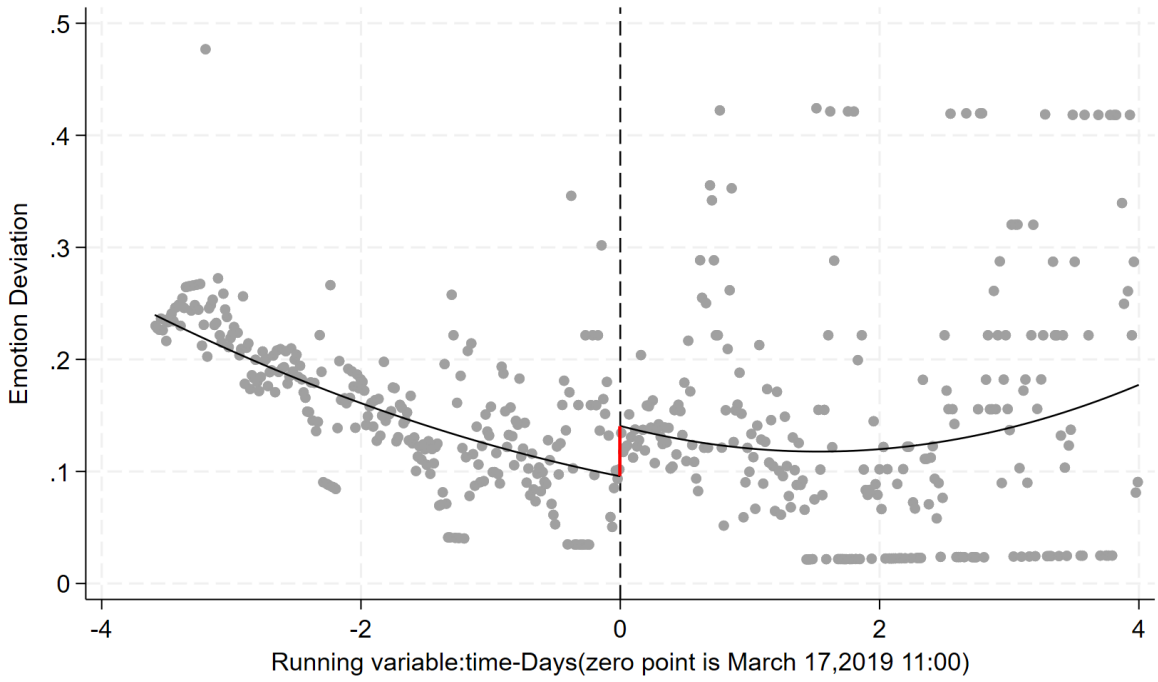

Figure 6 The Impact of Government Rumor Rebuttal on Emotion Deviation

Note The fitting line represents a quadratic fit of the emotion deviation to the running variable, with covariates controlled.

Figure Captions

Figure 1 Diagram of Discontinuity Design

Figure 2 Volume of Daily Posts on the Chengdu No.7 Middle School Food Safety Incident

Figure 3 The Impact of Government Rumor Rebuttal on Activists' Political Attitude, Factual Judgement and Emotion

Figure 4 The Impact of Government Rumor Rebuttal on Activists' Dissemination Behavior

Figure 5 The Impact of Government Rumor Rebuttal on the Dissidents' Discourse Deviation

Figure 6 The Impact of Government Rumor Rebuttal on Emotion Deviation