

# Reverse Engineering Chinese Censorship<sup>1</sup>

---

Gary King<sup>2</sup>

Institute for Quantitative Social Science  
Harvard University

(Talk at ESRC Research Methods Festival, 8 July 2014)

---

<sup>1</sup>Based on joint work with Jennifer Pan and Margaret Roberts

<sup>2</sup>GaryKing.org

- An Observational Study:

*How Censorship in China Allows Government Criticism but Silences Collective Expression* (*American Political Science Review*, May 2013)

- Experimental and Participatory Studies:

*Reverse Engineering Chinese Censorship* (*Under review*)

Copies at [GaryKing.org](http://GaryKing.org)

# Chinese Censorship

# Chinese Censorship

- The largest selective suppression of human expression in history:

# Chinese Censorship

- The largest selective suppression of human expression in history:
  - implemented *manually* (within a few hours of posting),

# Chinese Censorship

- The largest selective suppression of human expression in history:
  - implemented *manually* (within a few hours of posting),
  - by  $\approx 200,000$  workers,

# Chinese Censorship

- The largest selective suppression of human expression in history:
  - implemented *manually* (within a few hours of posting),
  - by  $\approx 200,000$  workers,
  - located in government and inside social media firms

# Chinese Censorship

- The largest selective suppression of human expression in history:
  - implemented *manually* (within a few hours of posting),
  - by  $\approx 200,000$  workers,
  - located in government and inside social media firms
- A huge censorship organization:

# Chinese Censorship

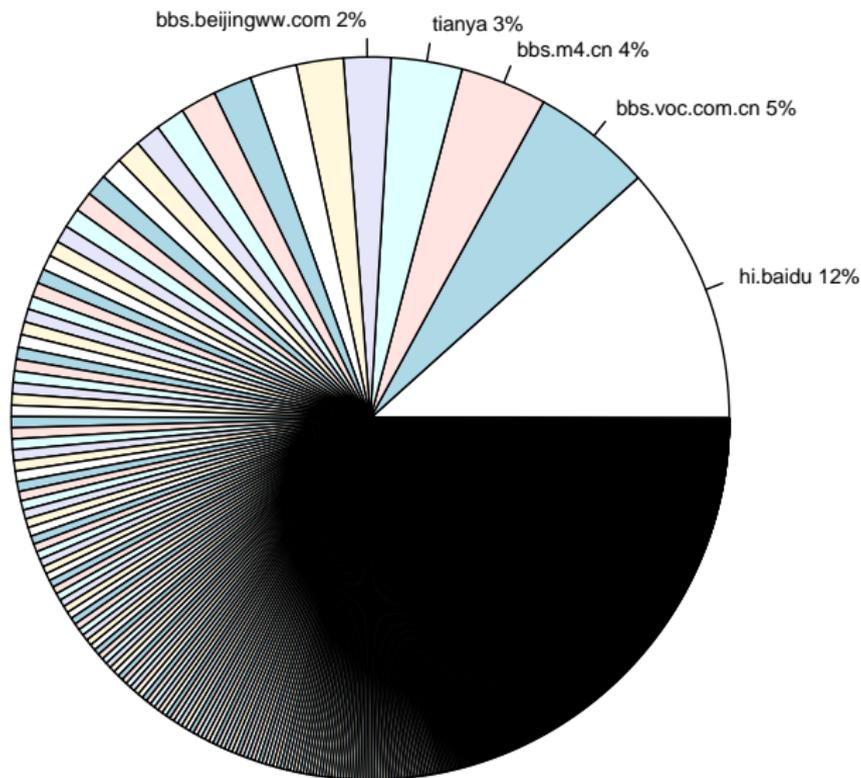
- The largest selective suppression of human expression in history:
  - implemented *manually* (within a few hours of posting),
  - by  $\approx 200,000$  workers,
  - located in government and inside social media firms
- A huge censorship organization:
  - (obviously) designed to suppress information

# Chinese Censorship

- The largest selective suppression of human expression in history:
  - implemented *manually* (within a few hours of posting),
  - by  $\approx 200,000$  workers,
  - located in government and inside social media firms
- A huge censorship organization:
  - (obviously) designed to suppress information
  - (paradoxically) very revealing about the goals, intentions, and actions of the Chinese leadership



# Chinese Social Media: Fractured over 1,400+ sites



(Plus 59% from Sina Blog)

# The Goals of Censorship make Social Media Actionable

# The Goals of Censorship make Social Media Actionable

- Everyone knows the Goal:

# The Goals of Censorship make Social Media Actionable

- Everyone knows the Goal:  
Stop criticism, protest, & collective action about the state,  
its leaders, and their policies

# The Goals of Censorship make Social Media Actionable

- ~~Everyone knows the Goal:~~  
~~Stop criticism, protest, & collective action about the state,~~  
~~its leaders, and their policies~~ *Wrong*

# The Goals of Censorship make Social Media Actionable

- ~~Everyone knows the Goal:~~  
Stop criticism, protest, & collective action about the state,  
its leaders, and their policies *Wrong*
- What Could be the Goal?

# The Goals of Censorship make Social Media Actionable

- ~~Everyone knows the Goal:~~  
~~Stop criticism, protest, & collective action about the state,~~  
~~its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  - 1 Stop criticism of the state

# The Goals of Censorship make Social Media Actionable

- ~~Everyone knows the Goal:~~  
~~Stop criticism, protest, & collective action about the state, its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  - 1 Stop criticism of the state
  - 2 Stop collective action

# The Goals of Censorship make Social Media Actionable

- ~~Everyone knows the Goal:~~  
~~Stop criticism, protest, & collective action about the state, its leaders, and their policies~~ *Wrong*
- ~~What Could be the Goal?~~
  - 1 ~~Stop criticism of the state~~ *Wrong*
  - 2 ~~Stop collective action~~

# The Goals of Censorship make Social Media Actionable

- ~~Everyone knows the Goal:~~  
~~Stop criticism, protest, & collective action about the state, its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  - 1 Stop criticism of the state *Wrong*
  - 2 Stop collective action *Right*

# The Goals of Censorship make Social Media Actionable

- ~~Everyone knows the Goal:~~  
~~Stop criticism, protest, & collective action about the state, its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  - ① ~~Stop criticism of the state~~ *Wrong*
  - ② Stop collective action *Right*
- Implications: Social Media is Actionable!

# The Goals of Censorship make Social Media Actionable

- ~~Everyone knows the Goal:~~  
~~Stop criticism, protest, & collective action about the state, its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  - ① ~~Stop criticism of the state~~ *Wrong*
  - ② Stop collective action *Right*
- Implications: Social Media is Actionable!
  - Chinese leaders:

# The Goals of Censorship make Social Media Actionable

- ~~Everyone knows the Goal:~~  
~~Stop criticism, protest, & collective action about the state, its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  - ① ~~Stop criticism of the state~~ *Wrong*
  - ② Stop collective action *Right*
- Implications: Social Media is Actionable!
  - Chinese leaders:
    - measure criticism: to judge local officials

# The Goals of Censorship make Social Media Actionable

- ~~Everyone knows the Goal:~~  
~~Stop criticism, protest, & collective action about the state, its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  - ① ~~Stop criticism of the state~~ *Wrong*
  - ② Stop collective action *Right*
- Implications: Social Media is Actionable!
  - Chinese leaders:
    - measure criticism: to judge local officials
    - censor: to stop events with collective action potential

# The Goals of Censorship make Social Media Actionable

- ~~Everyone knows the Goal:~~  
~~Stop criticism, protest, & collective action about the state, its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  - ① ~~Stop criticism of the state~~ *Wrong*
  - ② Stop collective action *Right*
- Implications: Social Media is Actionable!
  - Chinese leaders:
    - measure criticism: to judge local officials
    - censor: to stop events with collective action potential
  - Thus, we can use criticism & censorship to predict:

# The Goals of Censorship make Social Media Actionable

- ~~Everyone knows the Goal:~~  
~~Stop criticism, protest, & collective action about the state, its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  - ① ~~Stop criticism of the state~~ *Wrong*
  - ② Stop collective action *Right*
- Implications: Social Media is Actionable!
  - Chinese leaders:
    - measure criticism: to judge local officials
    - censor: to stop events with collective action potential
  - Thus, we can use criticism & censorship to predict:
    - Officials in trouble, likely to be replaced

# The Goals of Censorship make Social Media Actionable

- ~~Everyone knows the Goal:~~  
~~Stop criticism, protest, & collective action about the state, its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  - ① ~~Stop criticism of the state~~ *Wrong*
  - ② Stop collective action *Right*
- Implications: Social Media is Actionable!
  - Chinese leaders:
    - measure criticism: to judge local officials
    - censor: to stop events with collective action potential
  - Thus, we can use criticism & censorship to predict:
    - Officials in trouble, likely to be replaced
    - Policies that generate dissent (& interest of leaders)

# The Goals of Censorship make Social Media Actionable

- ~~Everyone knows the Goal:~~  
~~Stop criticism, protest, & collective action about the state, its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  - ① ~~Stop criticism of the state~~ *Wrong*
  - ② Stop collective action *Right*
- Implications: Social Media is Actionable!
  - Chinese leaders:
    - measure criticism: to judge local officials
    - censor: to stop events with collective action potential
  - Thus, we can use criticism & censorship to predict:
    - Officials in trouble, likely to be replaced
    - Policies that generate dissent (& interest of leaders)
    - Government action outside the Internet

# The Goals of Censorship make Social Media Actionable

- ~~Everyone knows the Goal:~~  
~~Stop criticism, protest, & collective action about the state, its leaders, and their policies~~ *Wrong*
- What Could be the Goal?
  - ① ~~Stop criticism of the state~~ *Wrong*
  - ② Stop collective action *Right*
- Implications: Social Media is Actionable!
  - Chinese leaders:
    - measure criticism: to judge local officials
    - censor: to stop events with collective action potential
  - Thus, we can use criticism & censorship to predict:
    - Officials in trouble, likely to be replaced
    - Policies that generate dissent (& interest of leaders)
    - Government action outside the Internet
    - Dissidents to be arrested; peace treaties to sign; emerging scandals

# The Goals of Censorship make Social Media Actionable

- ~~Everyone knows the Goal:~~  
~~Stop criticism, protest, & collective action about the state, its leaders, and their policies~~ *Wrong*
- **What Could be the Goal?**
  - ① ~~Stop criticism of the state~~ *Wrong*
  - ② ~~Stop collective action~~ *Right*
- **Implications: Social Media is Actionable!**
  - Chinese leaders:
    - measure criticism: to judge local officials
    - censor: to stop events with collective action potential
  - Thus, we can use criticism & censorship to predict:
    - Officials in trouble, likely to be replaced
    - Policies that generate dissent (& interest of leaders)
    - Government action outside the Internet
    - Dissidents to be arrested; peace treaties to sign; emerging scandals
    - Disagreements between central and local leaders

# The Goals of Censorship make Social Media Actionable

- ~~Everyone knows the Goal:~~  
~~Stop criticism, protest, & collective action about the state, its leaders, and their policies~~ *Wrong*
- ~~What Could be the Goal?~~
  - ① ~~Stop criticism of the state~~ *Wrong*
  - ② ~~Stop collective action~~ *Right*
- **Implications: Social Media is Actionable!**
  - Chinese leaders:
    - measure criticism: to judge local officials
    - censor: to stop events with collective action potential
  - Thus, we can use criticism & censorship to predict:
    - Officials in trouble, likely to be replaced
    - Policies that generate dissent (& interest of leaders)
    - Government action outside the Internet
    - Dissidents to be arrested; peace treaties to sign; emerging scandals
    - Disagreements between central and local leaders
- ↪ A big prize, but payable only if we have the goals right

# Our Observational Research Design: 5 Steps

# Our Observational Research Design: 5 Steps

- 1 **Monitor** 85 content areas over 6 months

# Our Observational Research Design: 5 Steps

- 1 **Monitor** 85 content areas over 6 months
- 2 **Download** each social media post, the instant it appears

# Our Observational Research Design: 5 Steps

- 1 **Monitor** 85 content areas over 6 months
- 2 **Download** each social media post, the instant it appears
- 3 **Revisit** each later to check if it was censored  
(carefully, from computers all over the world)

# Our Observational Research Design: 5 Steps

- 1 **Monitor** 85 content areas over 6 months
- 2 **Download** each social media post, the instant it appears
- 3 **Revisit** each later to check if it was censored  
(carefully, from computers all over the world)
- 4 **Repeat** 11,382,221 times

# Our Observational Research Design: 5 Steps

- 1 **Monitor** 85 content areas over 6 months
- 2 **Download** each social media post, the instant it appears
- 3 **Revisit** each later to check if it was censored  
(carefully, from computers all over the world)
- 4 **Repeat** 11,382,221 times
- 5 **Analyze** with new methods of computer-assisted text analysis

# Censorship is not Ambiguous: Example Error Page



The screenshot shows the top navigation bar of the 'Golden Sheep Community' (金羊社区) forum, including a search bar and a navigation menu. A central error message box is highlighted with a yellow border and contains the text: '抱歉, 指定的主题不存在或已被删除或正在被审核' (Sorry, the specified topic does not exist or has been deleted or is being investigated) and a link: '【点击这里返回上一页】' (Click here to return to the previous page). A red arrow points from the English text below to this message box.

Sorry, the host you were looking for does not exist, has been deleted, or is being investigated

# Censorship is not Ambiguous: BBS Error Page

# 404 ERROR

The page you requested is temporarily down. How about you go look at another page.



你访问的页面暂时找不到了哦，去看看别的页面吧。

[返回首页](#)

[反馈错误](#)

Jingjing, one of China's cartoon internet police

[关于我们](#) | [主页制作](#) | [客户服务](#) | [人才招聘](#) | [信息管理](#) | [业务联系](#) | [有奖新闻](#) | [网站地图](#)

Copyright(c) (2007-2012)New Silkroad Online. All rights reserved.

[新ICP备07500354号]互联网违法和不良信息举报中心 有害信息举报中心 互动频道举报奖励办法

版权所有：中国电信股份有限公司新疆分公司 技术支持：800-8930169 电话：0991-4667780 传真：(0991)4682953

电子信箱：[edit@mail.xj.cninfo.net](mailto:edit@mail.xj.cninfo.net) 文化部网络游戏运营许可证 编号：文网文(2010)054号

增值电信业务经营许可证A2.B1.B2-20090001 网络文化经营许可证080626 广播电视节目制作经营许可证编号：(新)字第051号



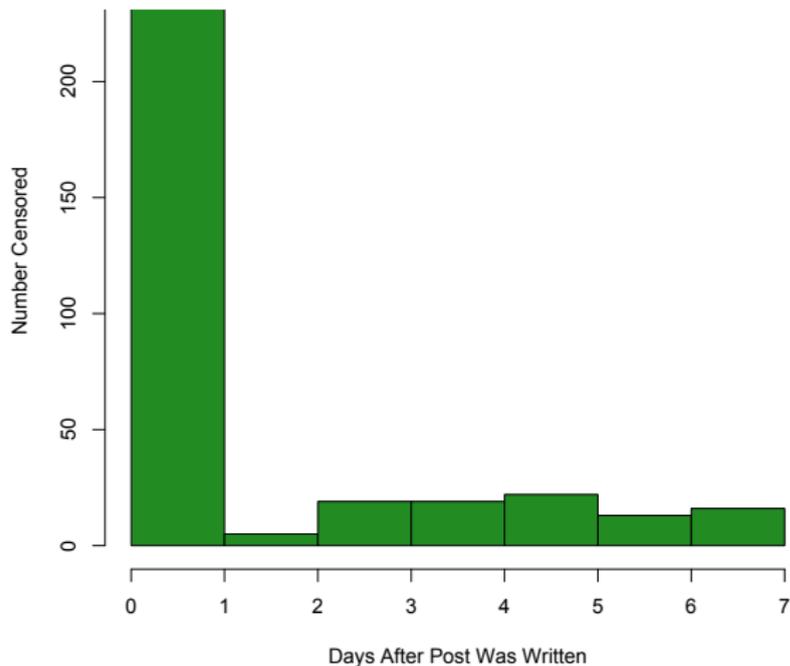
The Censors are Fast; Our Automated Methods are Faster

# The Censors are Fast; Our Automated Methods are Faster

Example: Shanghai Subway Crash

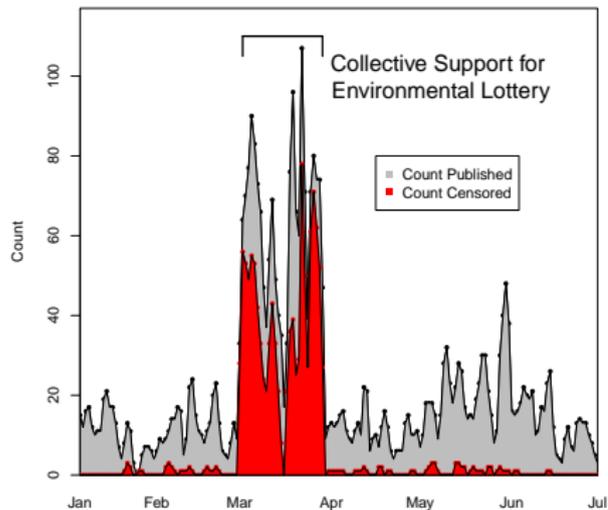
# The Censors are Fast; Our Automated Methods are Faster

Example: Shanghai Subway Crash

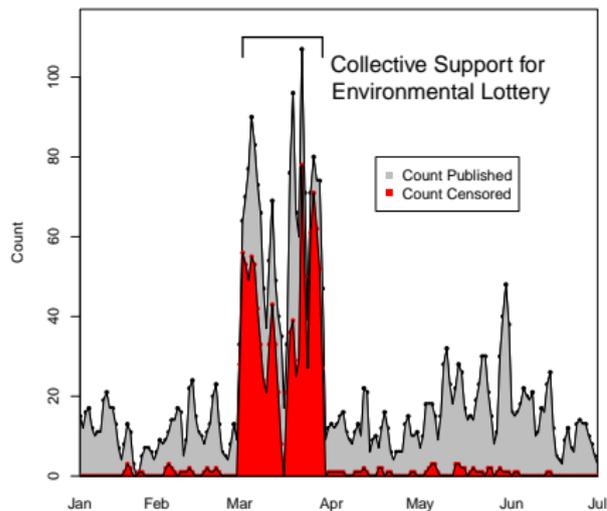


## Post Volume & Censorship are “Bursty”

# Post Volume & Censorship are “Bursty”

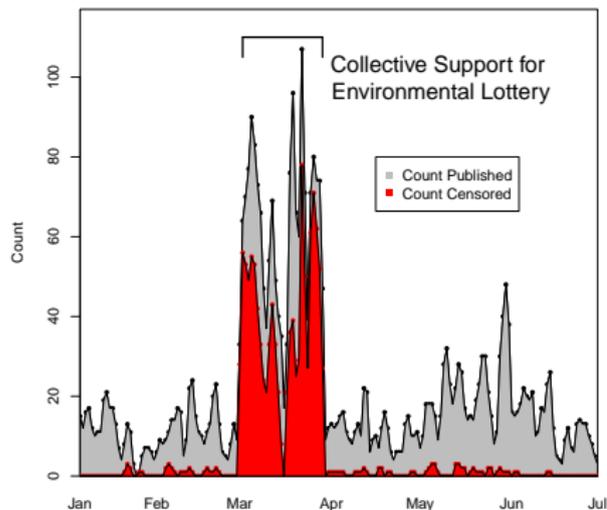


## Post Volume & Censorship are “Bursty”



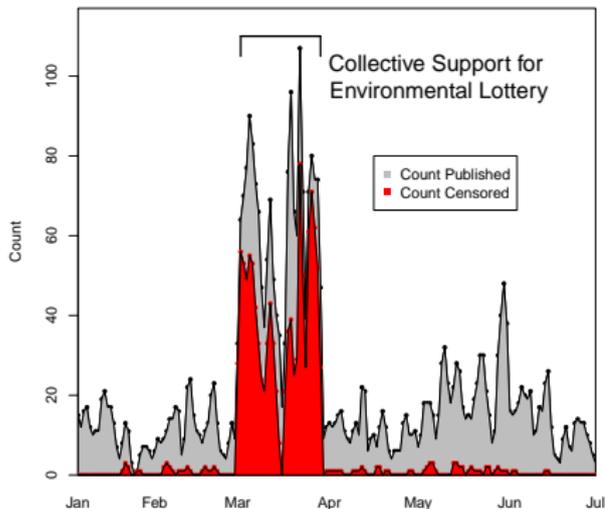
- Found 87 volume bursts in our 85 topic areas

## Post Volume & Censorship are “Bursty”



- Found 87 volume bursts in our 85 topic areas
- Identified real world event associated with each burst

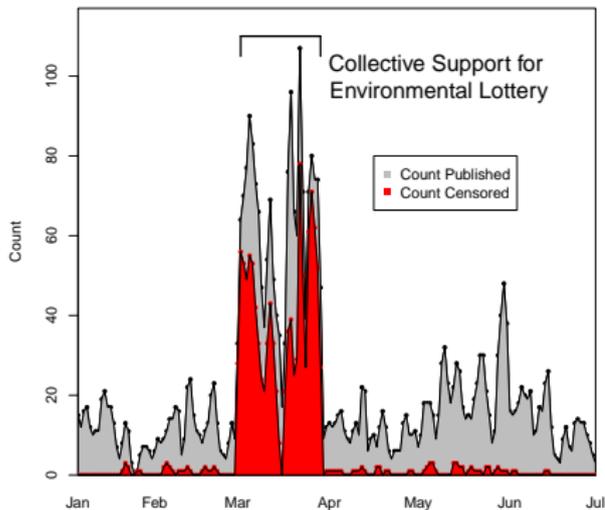
## Post Volume & Censorship are “Bursty”



- Found 87 volume bursts in our 85 topic areas
- Identified real world event associated with each burst

Our hypothesis: The government censors all posts in volume bursts associated with events with collective action potential (regardless of how critical or supportive of the state)

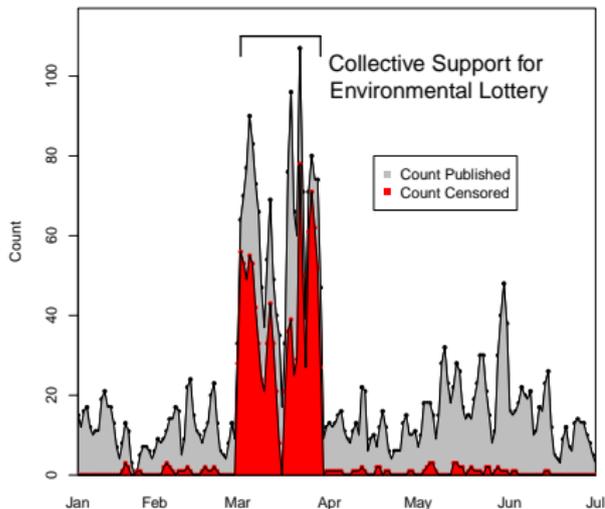
## Post Volume & Censorship are “Bursty”



- Found 87 volume bursts in our 85 topic areas
- Identified real world event associated with each burst

Our hypothesis: The government censors all posts in volume bursts associated with events with collective action potential (regardless of how critical or supportive of the state)

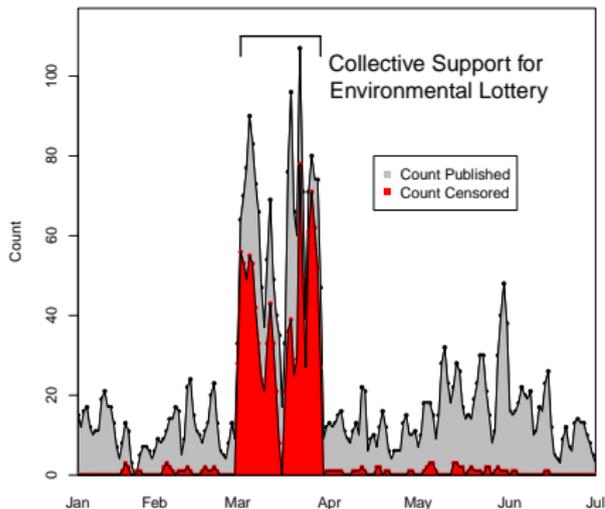
## Post Volume & Censorship are “Bursty”



- Found 87 volume bursts in our 85 topic areas
- Identified real world event associated with each burst

Our hypothesis: The government censors all posts in volume bursts associated with events with collective action potential (regardless of how critical or supportive of the state)

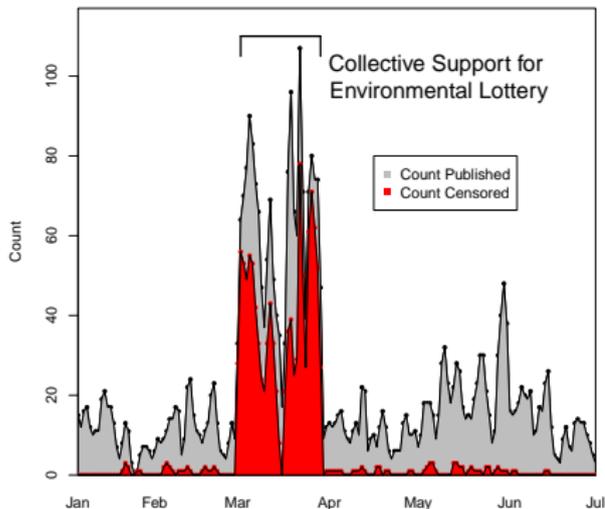
## Post Volume & Censorship are “Bursty”



- Found 87 volume bursts in our 85 topic areas
- Identified real world event associated with each burst

Our hypothesis: The government censors all posts in volume bursts associated with events with collective action potential (regardless of how critical or supportive of the state)

## Post Volume & Censorship are “Bursty”



- Found 87 volume bursts in our 85 topic areas
- Identified real world event associated with each burst

Our hypothesis: The government censors all posts in volume bursts associated with events with collective action potential (regardless of how critical or supportive of the state)

# Observational Test 1: Post Volume

## Observational Test 1: Post Volume

- Begin with our 87 volume bursts in 85 topics areas

## Observational Test 1: Post Volume

- Begin with our 87 volume bursts in 85 topics areas
- For each burst, calculate change in % censorship inside to outside each volume burst within topic areas – censorship magnitude

## Observational Test 1: Post Volume

- Begin with our 87 volume bursts in 85 topics areas
- For each burst, calculate change in % censorship inside to outside each volume burst within topic areas – censorship magnitude
- If goal of censorship is to stop collective action, we expect:

## Observational Test 1: Post Volume

- Begin with our 87 volume bursts in 85 topics areas
  - For each burst, calculate change in % censorship inside to outside each volume burst within topic areas – censorship magnitude
  - If goal of censorship is to stop collective action, we expect:
- 1 On average, % censored should increase during volume bursts

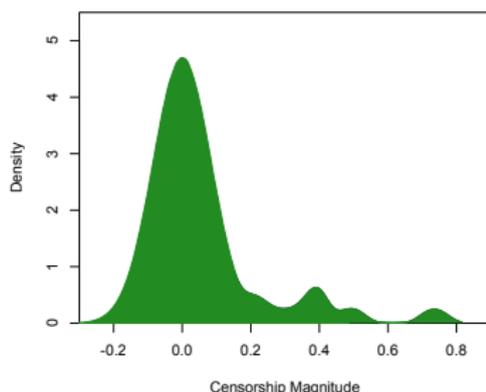
## Observational Test 1: Post Volume

- Begin with our **87 volume bursts** in 85 topics areas
- For each burst, calculate change in % censorship inside to outside each volume burst within topic areas – **censorship magnitude**
- If goal of censorship is to stop collective action, **we expect:**
  - 1 On average, % censored should increase during volume bursts
  - 2 Some bursts (associated with politically relevant events) should have much higher censorship

# Observational Test 1: Post Volume

- Begin with our **87 volume bursts** in 85 topics areas
- For each burst, calculate change in % censorship inside to outside each volume burst within topic areas – **censorship magnitude**
- If goal of censorship is to stop collective action, **we expect:**

- 1 On average, % censored should increase during volume bursts
- 2 Some bursts (associated with politically relevant events) should have much higher censorship



# Classification of Events Generating the Bursts

## Classification of Events Generating the Bursts

Event classification (each category can be +, -, or neutral comments about the state)

# Classification of Events Generating the Bursts

Event classification (each category can be +, -, or neutral comments about the state)

- 1 Collective Action Potential

# Classification of Events Generating the Bursts

Event classification (each category can be +, -, or neutral comments about the state)

- ① **Collective Action Potential**
  - protest or organized crowd formation outside the Internet

# Classification of Events Generating the Bursts

Event classification (each category can be +, -, or neutral comments about the state)

## ① Collective Action Potential

- protest or organized crowd formation outside the Internet
- individuals who have organized or incited collective action on the ground in the past;

# Classification of Events Generating the Bursts

Event classification (each category can be +, -, or neutral comments about the state)

## ① Collective Action Potential

- protest or organized crowd formation outside the Internet
- individuals who have organized or incited collective action on the ground in the past;
- topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.

# Classification of Events Generating the Bursts

Event classification (each category can be +, -, or neutral comments about the state)

## ① Collective Action Potential

- protest or organized crowd formation outside the Internet
- individuals who have organized or incited collective action on the ground in the past;
- topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
- (Inter-coder reliability: 98.9%)

# Classification of Events Generating the Bursts

Event classification (each category can be +, -, or neutral comments about the state)

## ① Collective Action Potential

- protest or organized crowd formation outside the Internet
- individuals who have organized or incited collective action on the ground in the past;
- topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
- (Inter-coder reliability: 98.9%)

## ② Criticism of censors

# Classification of Events Generating the Bursts

Event classification (each category can be +, -, or neutral comments about the state)

## ① Collective Action Potential

- protest or organized crowd formation outside the Internet
- individuals who have organized or incited collective action on the ground in the past;
- topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
- (Inter-coder reliability: 98.9%)

## ② Criticism of censors

## ③ Pornography

# Classification of Events Generating the Bursts

Event classification (each category can be +, -, or neutral comments about the state)

## ① Collective Action Potential

- protest or organized crowd formation outside the Internet
- individuals who have organized or incited collective action on the ground in the past;
- topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
- (Inter-coder reliability: 98.9%)

## ② Criticism of censors

## ③ Pornography

## ④ (Other) News

# Classification of Events Generating the Bursts

Event classification (each category can be +, -, or neutral comments about the state)

## ① Collective Action Potential

- protest or organized crowd formation outside the Internet
- individuals who have organized or incited collective action on the ground in the past;
- topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
- (Inter-coder reliability: 98.9%)

## ② Criticism of censors

## ③ Pornography

## ④ (Other) News

## ⑤ Government Policies

# Classification of Events Generating the Bursts

Event classification (each category can be +, -, or neutral comments about the state)

- 1 ~~Collective Action Potential~~ 
  - protest or organized crowd formation outside the Internet
  - individuals who have organized or incited collective action on the ground in the past;
  - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
  - (Inter-coder reliability: 98.9%)
- 2 Criticism of censors
- 3 Pornography
- 4 (Other) News
- 5 Government Policies

# Classification of Events Generating the Bursts

Event classification (each category can be +, -, or neutral comments about the state)

- 1 ~~Collective Action Potential~~ **CENSORED**
  - protest or organized crowd formation outside the Internet
  - individuals who have organized or incited collective action on the ground in the past;
  - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
  - (Inter-coder reliability: 98.9%)
- 2 ~~Criticism of censors~~ **CENSORED**
- 3 Pornography
- 4 (Other) News
- 5 Government Policies

# Classification of Events Generating the Bursts

Event classification (each category can be +, -, or neutral comments about the state)

- 1 ~~Collective Action Potential~~ **CENSORED**
  - protest or organized crowd formation outside the Internet
  - individuals who have organized or incited collective action on the ground in the past;
  - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
  - (Inter-coder reliability: 98.9%)
- 2 ~~Criticism of censors~~ **CENSORED**
- 3 ~~Pornography~~ **CENSORED**
- 4 (Other) News
- 5 Government Policies

# Classification of Events Generating the Bursts

Event classification (each category can be +, -, or neutral comments about the state)

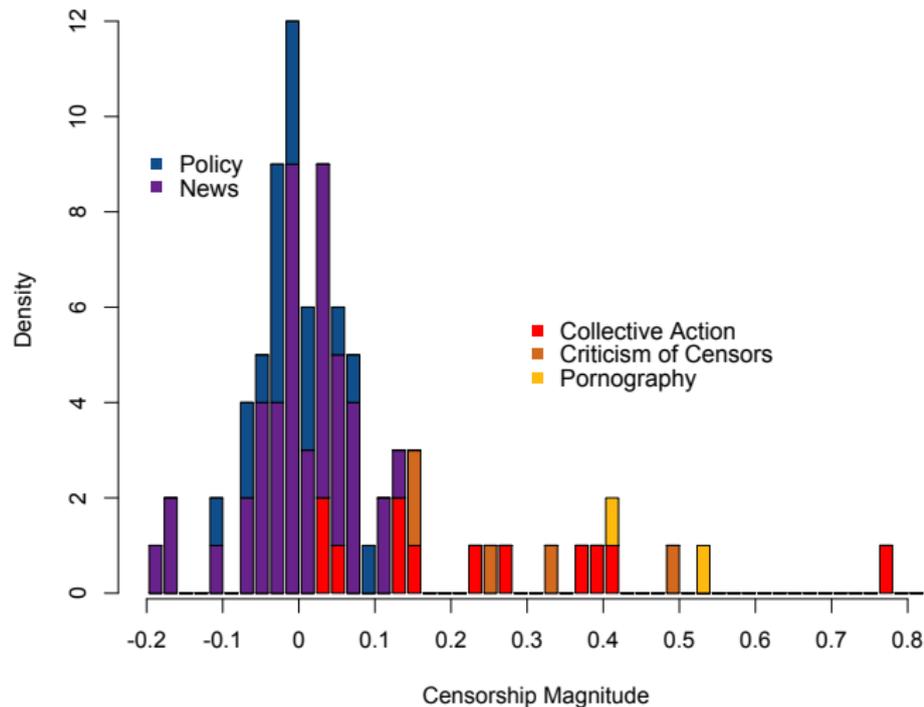
- 1 ~~Collective Action Potential~~ 
  - protest or organized crowd formation outside the Internet
  - individuals who have organized or incited collective action on the ground in the past;
  - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
  - (Inter-coder reliability: 98.9%)
- 2 ~~Criticism of censors~~ 
- 3 ~~Pornography~~ 
- 4 (Other) News 
- 5 Government Policies

# Classification of Events Generating the Bursts

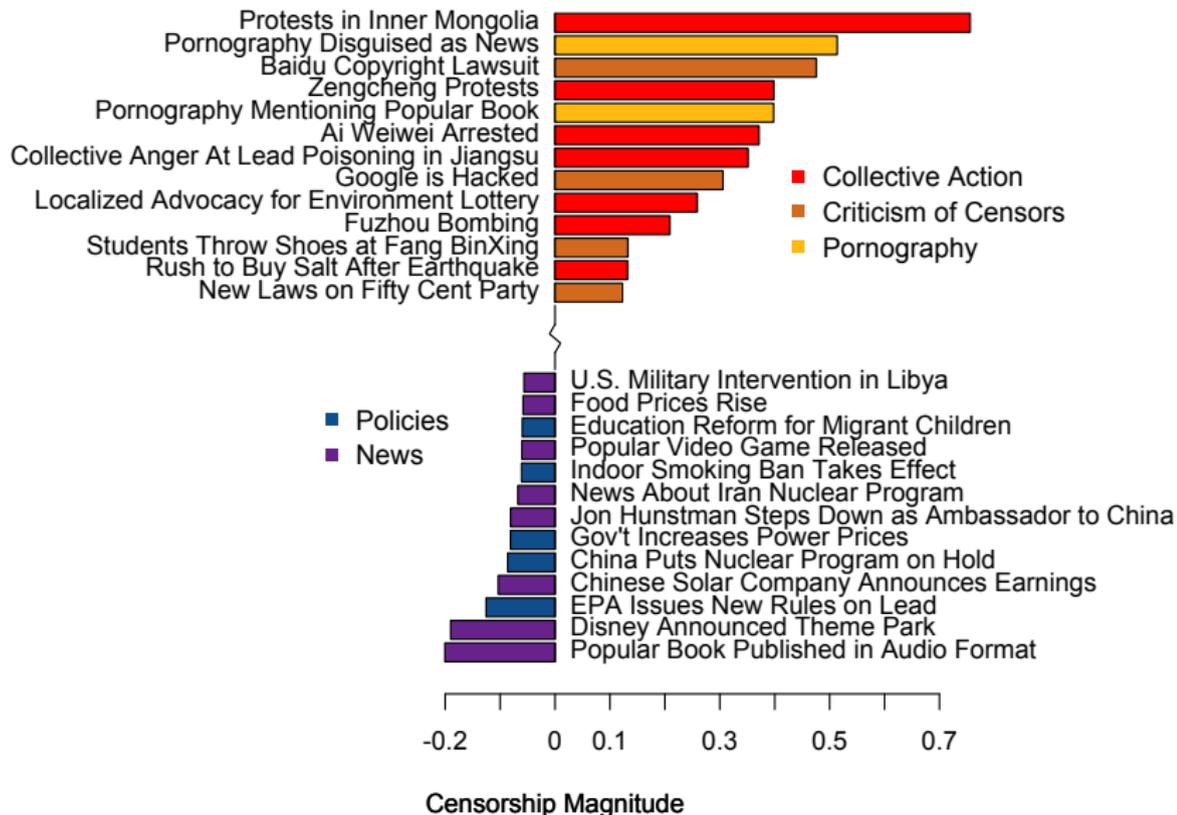
Event classification (each category can be +, -, or neutral comments about the state)

- 1 ~~Collective Action Potential~~ 
  - protest or organized crowd formation outside the Internet
  - individuals who have organized or incited collective action on the ground in the past;
  - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
  - (Inter-coder reliability: 98.9%)
- 2 ~~Criticism of censors~~ 
- 3 ~~Pornography~~ 
- 4 (Other) News 
- 5 Government Policies 

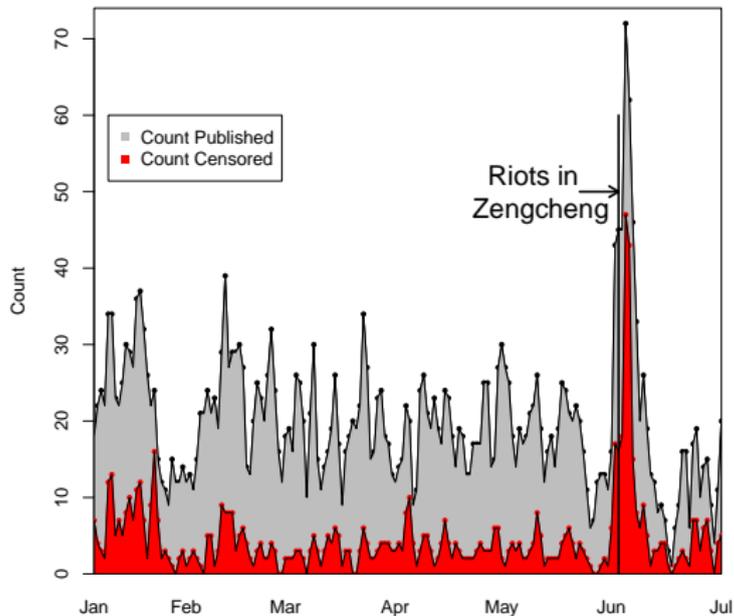
# What Types of Events Are Censored?



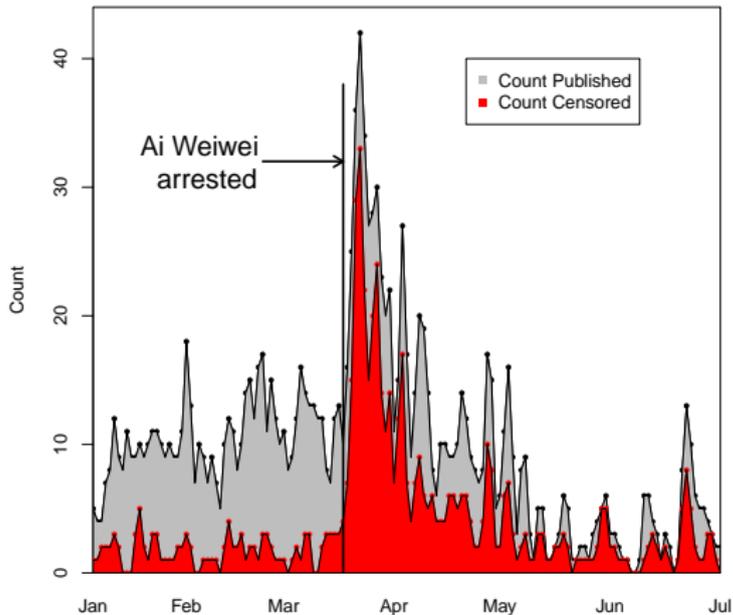
# What Types of Events Are Censored?



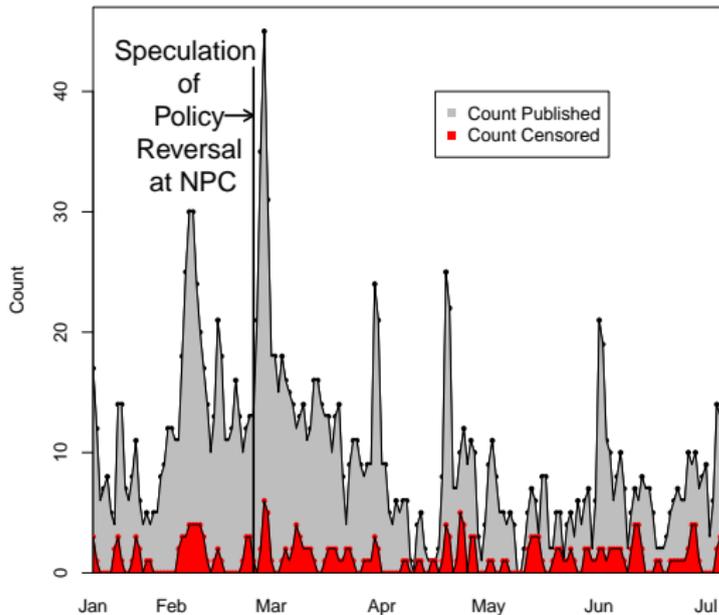
# Censoring Collective Action: Riots in Zengcheng



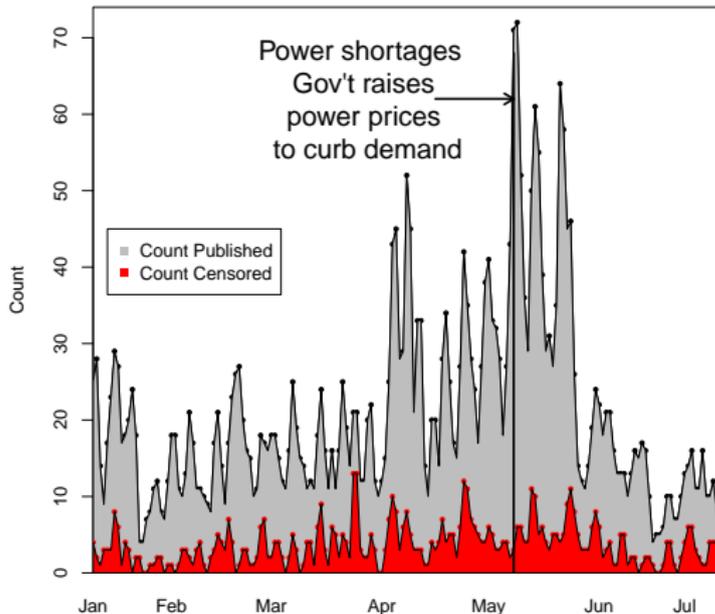
# Censoring Collective Action: Ai Weiwei's Arrest



# Low Censorship on Policy: One Child



# Low Censorship on News: Power Prices



# Content of Posts

## Content of Posts

We adapted “ReadMe” (Hopkins-King) methodology to the content of censored posts in Chinese

## Content of Posts

We adapted “ReadMe” (Hopkins-King) methodology to the content of censored posts in Chinese

- Standardize encoding of Chinese text

## Content of Posts

We adapted “ReadMe” (Hopkins-King) methodology to the content of censored posts in Chinese

- Standardize encoding of Chinese text
- Eliminate punctuation and stop words

## Content of Posts

We adapted “ReadMe” (Hopkins-King) methodology to the content of censored posts in Chinese

- Standardize encoding of Chinese text
- Eliminate punctuation and stop words
- Experiment with segmenting characters into words

## Content of Posts

We adapted “ReadMe” (Hopkins-King) methodology to the content of censored posts in Chinese

- Standardize encoding of Chinese text
- Eliminate punctuation and stop words
- Experiment with segmenting characters into words
- Calculate the quantity of interest (% censored by category)

## Content of Posts

We adapted “ReadMe” (Hopkins-King) methodology to the content of censored posts in Chinese

- Standardize encoding of Chinese text
- Eliminate punctuation and stop words
- Experiment with segmenting characters into words
- Calculate the quantity of interest (% censored by category)
  - Code a validation set

## Content of Posts

We adapted “ReadMe” (Hopkins-King) methodology to the content of censored posts in Chinese

- Standardize encoding of Chinese text
- Eliminate punctuation and stop words
- Experiment with segmenting characters into words
- Calculate the quantity of interest (% censored by category)
  - Code a validation set
  - ReadMe estimates: % of posts in each category

# Content of Posts

We adapted “ReadMe” (Hopkins-King) methodology to the content of censored posts in Chinese

- Standardize encoding of Chinese text
- Eliminate punctuation and stop words
- Experiment with segmenting characters into words
- Calculate the quantity of interest (% censored by category)
  - Code a validation set
  - ReadMe estimates: % of posts in each category
  - Quantity of interest: % of posts censored in a category

# Content of Posts

We adapted “ReadMe” (Hopkins-King) methodology to the content of censored posts in Chinese

- Standardize encoding of Chinese text
- Eliminate punctuation and stop words
- Experiment with segmenting characters into words
- Calculate the quantity of interest (% censored by category)
  - Code a validation set
  - ReadMe estimates: % of posts in each category
  - Quantity of interest: % of posts censored in a category

$$P(\text{Censored}|\text{Category}) = \frac{P(\text{Category}|\text{Censored})P(\text{Censored})}{P(\text{Category})}$$

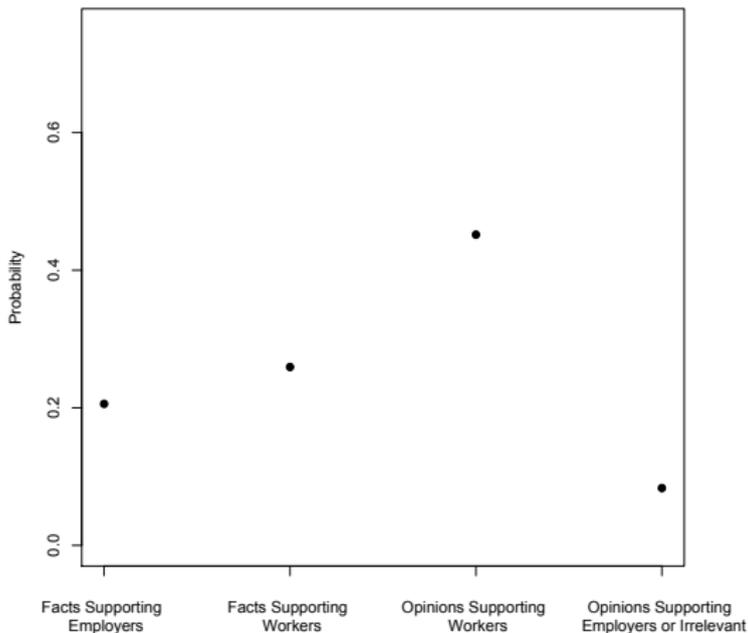
# “ReadMe” Algorithm Validated in Chinese

# “ReadMe” Algorithm Validated in Chinese

Example: Labor Strikes, 2010 (Training set: 100; Test set: 900)

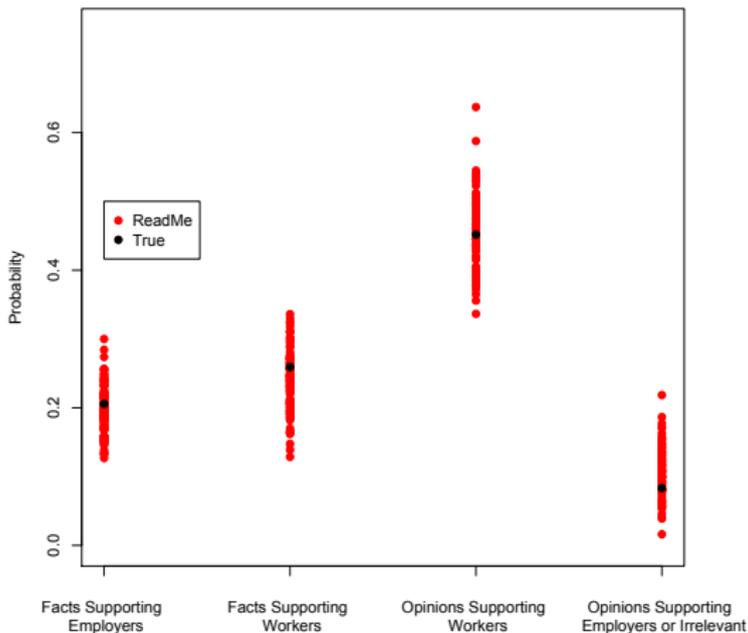
# “ReadMe” Algorithm Validated in Chinese

Example: Labor Strikes, 2010 (Training set: 100; Test set: 900)



# “ReadMe” Algorithm Validated in Chinese

Example: Labor Strikes, 2010 (Training set: 100; Test set: 900)



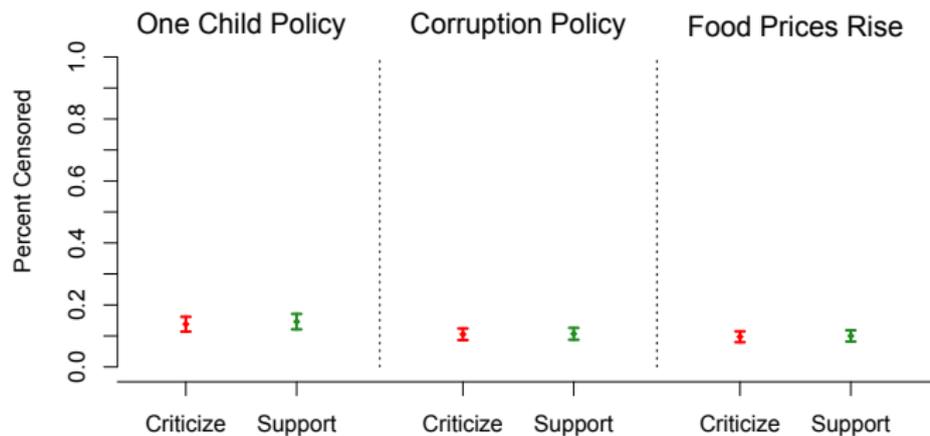
## Uncensored: Non-Collective Action Posts

# Uncensored: Non-Collective Action Posts

Regardless of whether Supportive *or* Critical

# Uncensored: Non-Collective Action Posts

Regardless of whether Supportive or Critical



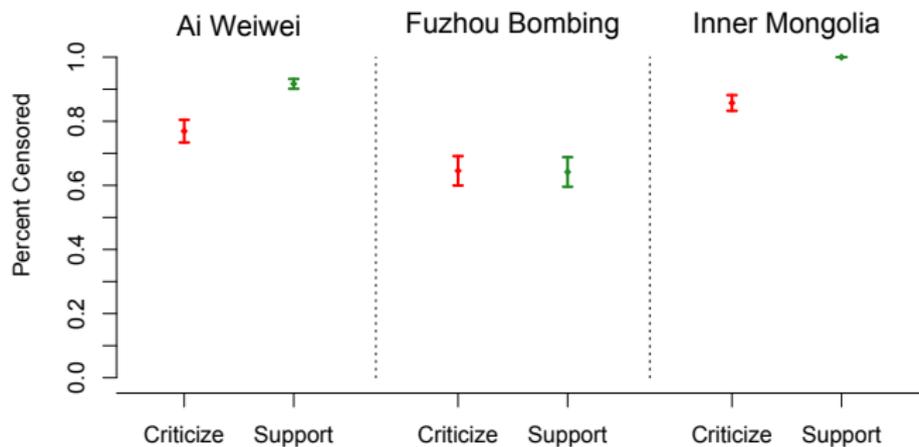
## Censored: Collective Action Posts

# Censored: Collective Action Posts

Regardless of whether Supportive *or* Critical

# Censored: Collective Action Posts

Regardless of whether Supportive or Critical



# Additional Research Designs

# Additional Research Designs

- 1 Randomized Experiment (for causal inferences)

# Additional Research Designs

- 1 Randomized Experiment (for causal inferences)
- 2 Participatory Study (for descriptive inferences)

# Additional Research Designs

- ① Randomized Experiment (for causal inferences)
  - Selected 100 top social media sites ( $\sim 87\%$  of blogs,  $> 500\text{M}$  Users, geographically diverse)
  
- ② Participatory Study (for descriptive inferences)

# Additional Research Designs

- ① Randomized Experiment (for causal inferences)
  - Selected 100 top social media sites (~87% of blogs, >500M Users, geographically diverse)
  - Created 2 accounts on each (from inside China all over the country)
  
- ② Participatory Study (for descriptive inferences)

# Additional Research Designs

- ① Randomized Experiment (for causal inferences)
  - Selected 100 top social media sites (~87% of blogs, >500M Users, geographically diverse)
  - Created 2 accounts on each (from inside China all over the country)
  - Wrote 1,200 *unique* social media posts (CA/not CA, Pro/Anti)
  
- ② Participatory Study (for descriptive inferences)

# Additional Research Designs

- 1 Randomized Experiment (for causal inferences)
  - Selected 100 top social media sites (~87% of blogs, >500M Users, geographically diverse)
  - Created 2 accounts on each (from inside China all over the country)
  - Wrote 1,200 *unique* social media posts (CA/not CA, Pro/Anti)
  - Submitted posts randomly assigned to type
  
- 2 Participatory Study (for descriptive inferences)

# Additional Research Designs

- 1 Randomized Experiment (for causal inferences)
  - Selected 100 top social media sites ( $\sim 87\%$  of blogs,  $> 500\text{M}$  Users, geographically diverse)
  - Created 2 accounts on each (from inside China all over the country)
  - Wrote 1,200 *unique* social media posts (CA/not CA, Pro/Anti)
  - Submitted posts randomly assigned to type
  - Checked on censorship (from computers in many countries)
- 2 Participatory Study (for descriptive inferences)

# Additional Research Designs

## ① Randomized Experiment (for causal inferences)

- Selected 100 top social media sites (~87% of blogs, >500M Users, geographically diverse)
- Created 2 accounts on each (from inside China all over the country)
- Wrote 1,200 *unique* social media posts (CA/not CA, Pro/Anti)
- Submitted posts randomly assigned to type
- Checked on censorship (from computers in many countries)

## ② Participatory Study (for descriptive inferences)

- Current method of learning how they censor: ask (carefully!)

# Additional Research Designs

## ① Randomized Experiment (for causal inferences)

- Selected 100 top social media sites (~87% of blogs, >500M Users, geographically diverse)
- Created 2 accounts on each (from inside China all over the country)
- Wrote 1,200 *unique* social media posts (CA/not CA, Pro/Anti)
- Submitted posts randomly assigned to type
- Checked on censorship (from computers in many countries)

## ② Participatory Study (for descriptive inferences)

- Current method of learning how they censor: ask (carefully!)
- Our goal: change our sources' incentives

# Additional Research Designs

## 1 Randomized Experiment (for causal inferences)

- Selected 100 top social media sites (~87% of blogs, >500M Users, geographically diverse)
- Created 2 accounts on each (from inside China all over the country)
- Wrote 1,200 *unique* social media posts (CA/not CA, Pro/Anti)
- Submitted posts randomly assigned to type
- Checked on censorship (from computers in many countries)

## 2 Participatory Study (for descriptive inferences)

- Current method of learning how they censor: ask (carefully!)
- Our goal: change our sources' incentives
- Procedure: create our own social media website in China

# Additional Research Designs

## 1 Randomized Experiment (for causal inferences)

- Selected 100 top social media sites (~87% of blogs, >500M Users, geographically diverse)
- Created 2 accounts on each (from inside China all over the country)
- Wrote 1,200 *unique* social media posts (CA/not CA, Pro/Anti)
- Submitted posts randomly assigned to type
- Checked on censorship (from computers in many countries)

## 2 Participatory Study (for descriptive inferences)

- Current method of learning how they censor: ask (carefully!)
- Our goal: change our sources' incentives
- Procedure: create our own social media website in China
- Bought URL; contracted with firms for servers & software; posted and censored ourselves

# Additional Research Designs

## ① Randomized Experiment (for causal inferences)

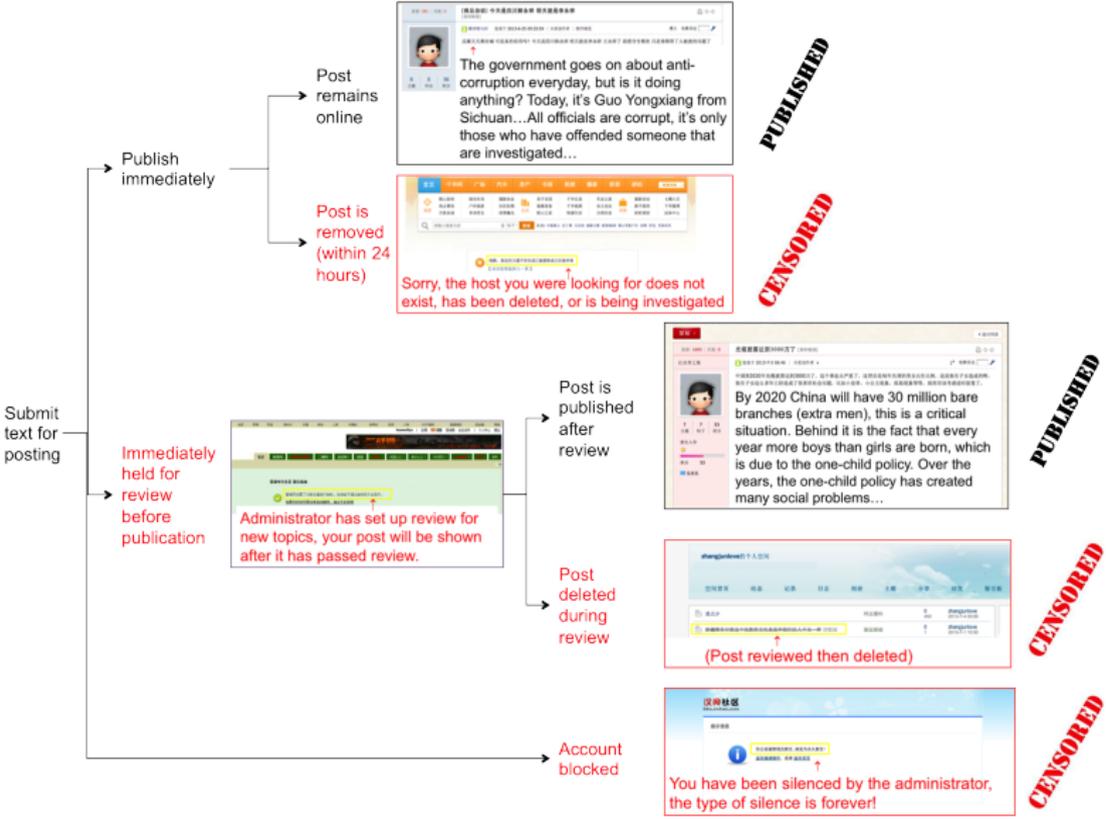
- Selected 100 top social media sites (~87% of blogs, >500M Users, geographically diverse)
- Created 2 accounts on each (from inside China all over the country)
- Wrote 1,200 *unique* social media posts (CA/not CA, Pro/Anti)
- Submitted posts randomly assigned to type
- Checked on censorship (from computers in many countries)

## ② Participatory Study (for descriptive inferences)

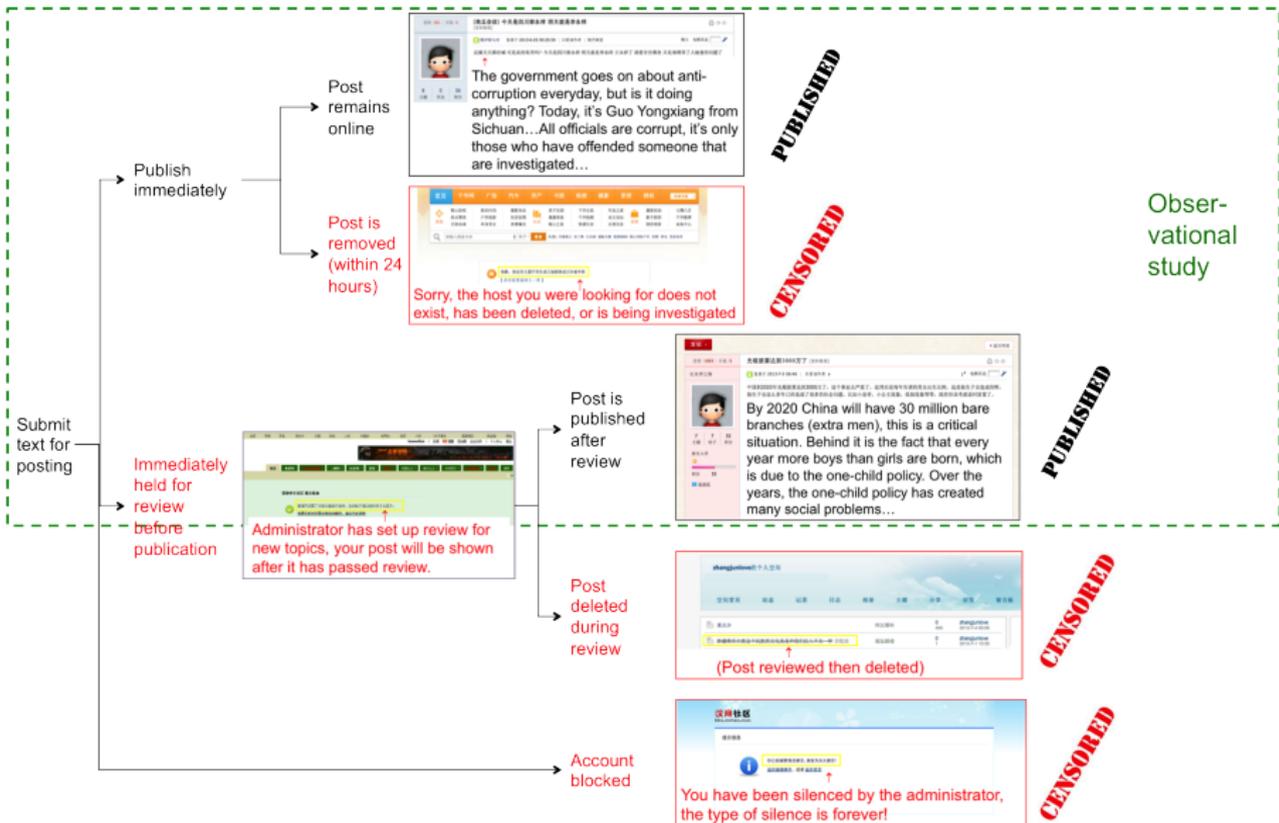
- Current method of learning how they censor: ask (carefully!)
- Our goal: change our sources' incentives
- Procedure: create our own social media website in China
- Bought URL; contracted with firms for servers & software; posted and censored ourselves
- To learn: we tried every software option, read the documentation, and called customer support(!)

# Mechanisms of Censorship

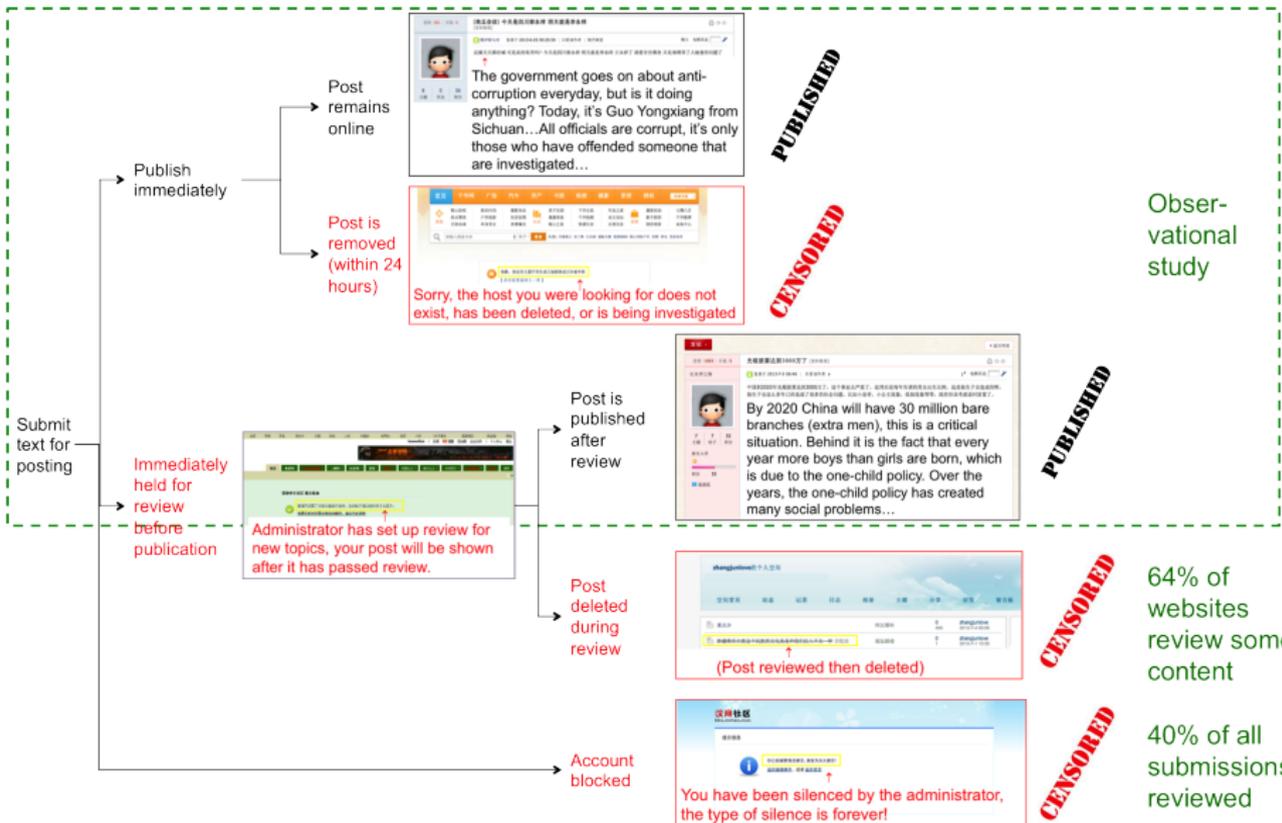
# Mechanisms of Censorship



# Mechanisms of Censorship

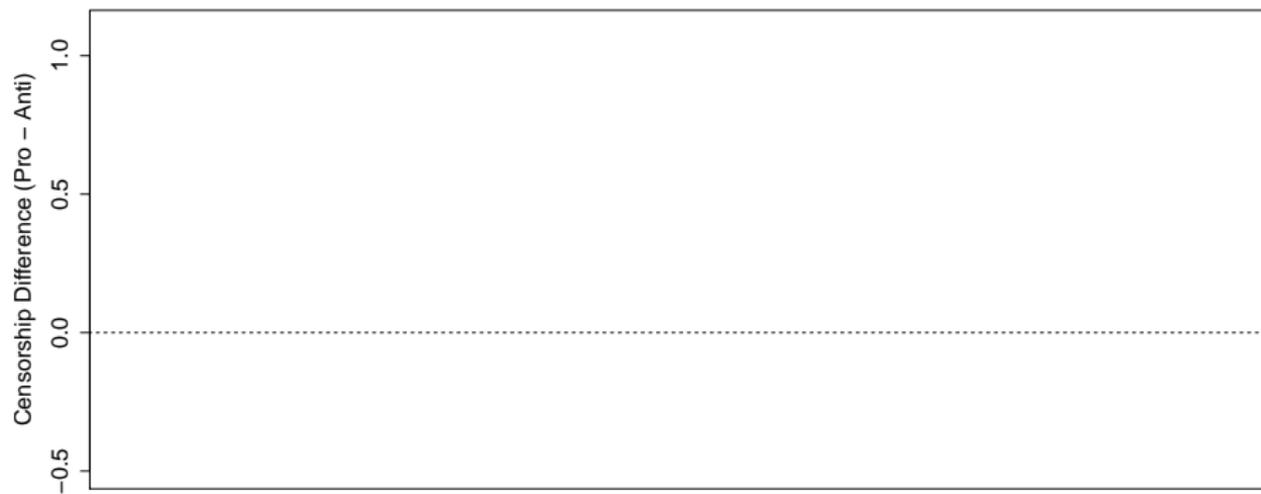


# Mechanisms of Censorship

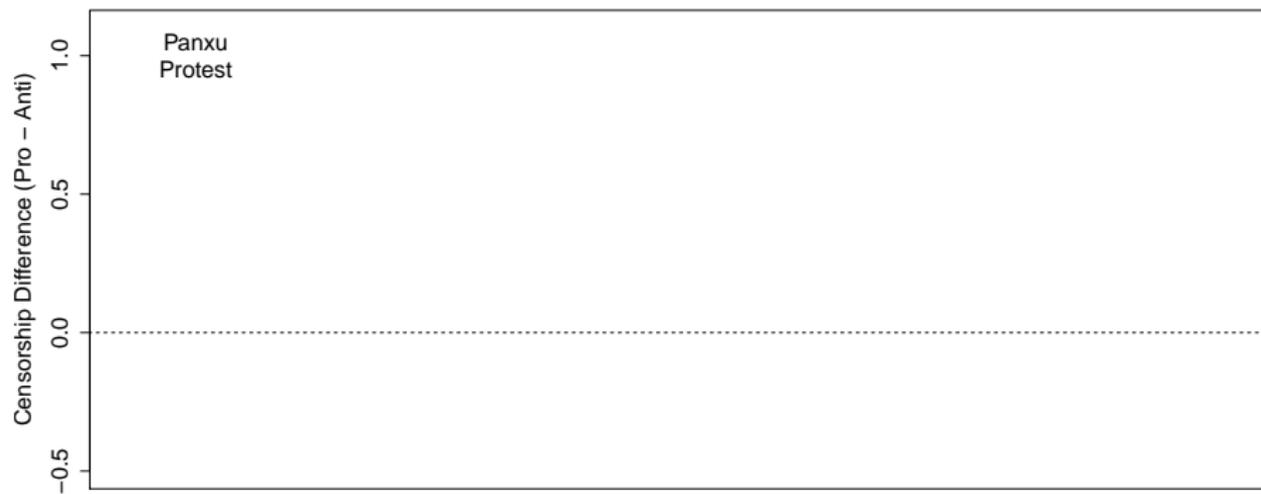


# Posts For v. Against Government: Zero Causal Effect

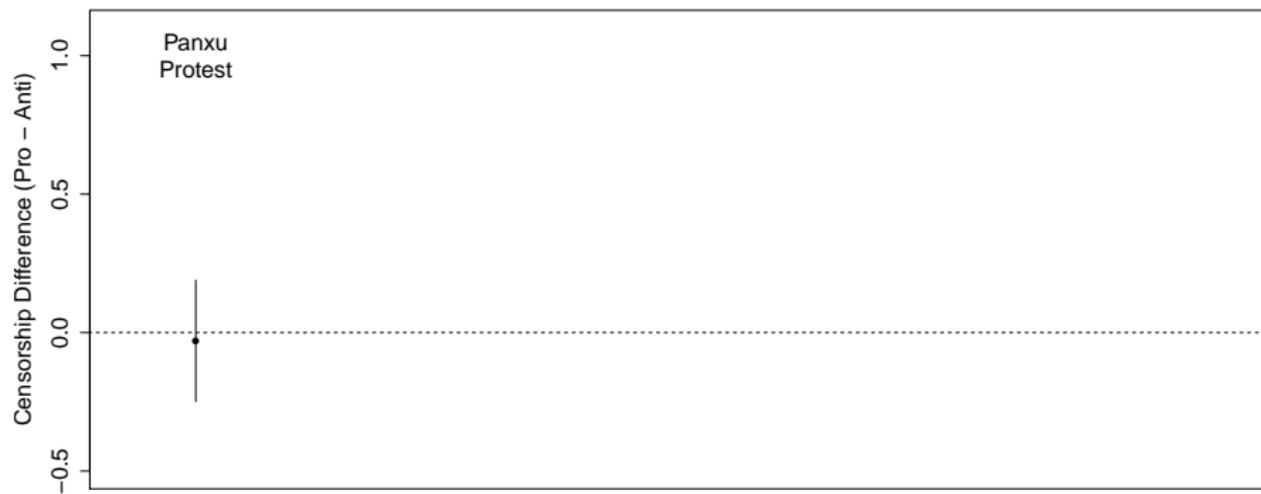
## Posts For v. Against Government: Zero Causal Effect



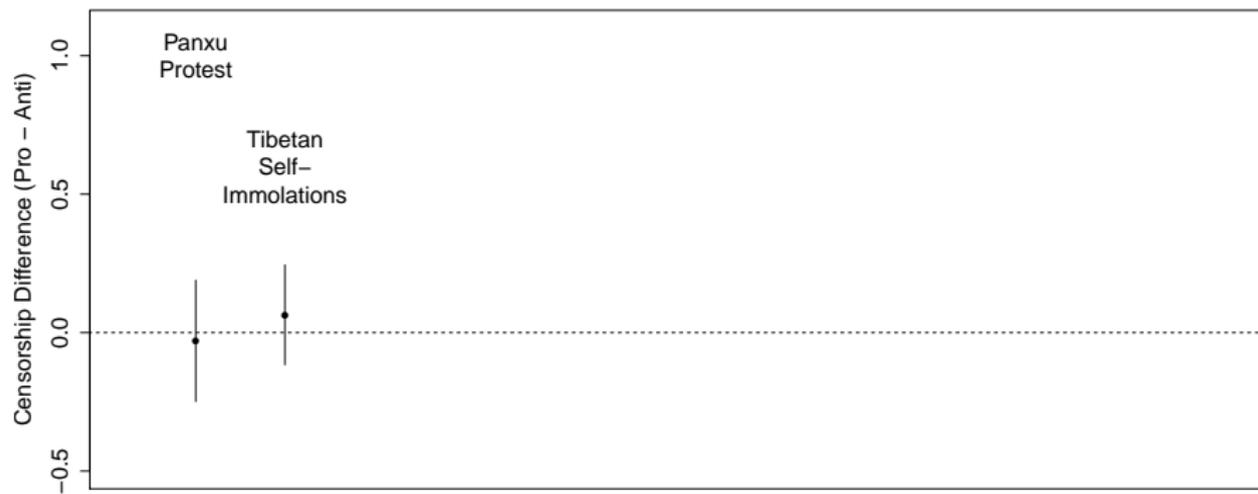
## Posts For v. Against Government: Zero Causal Effect



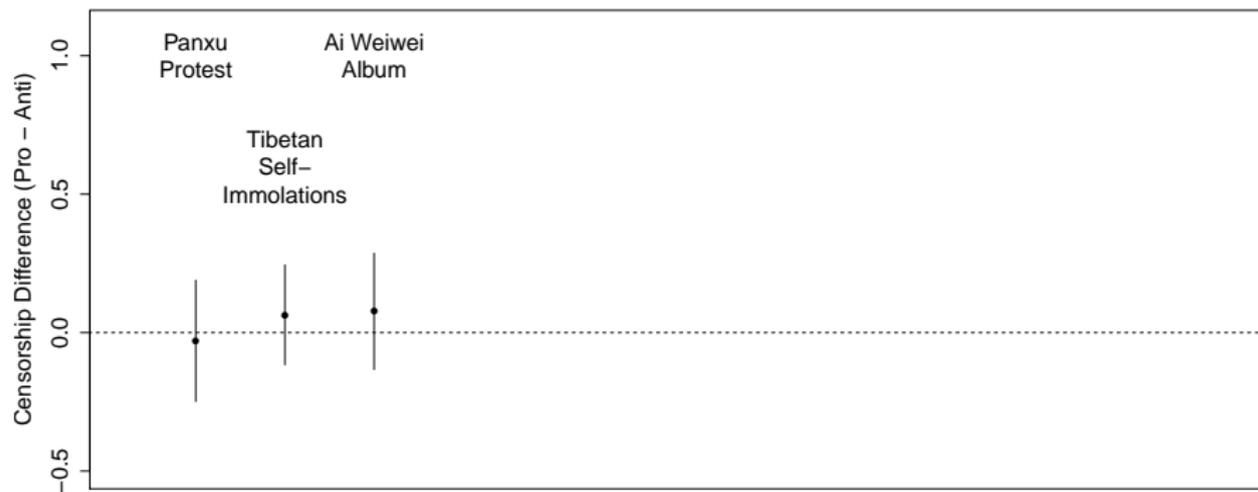
## Posts For v. Against Government: Zero Causal Effect



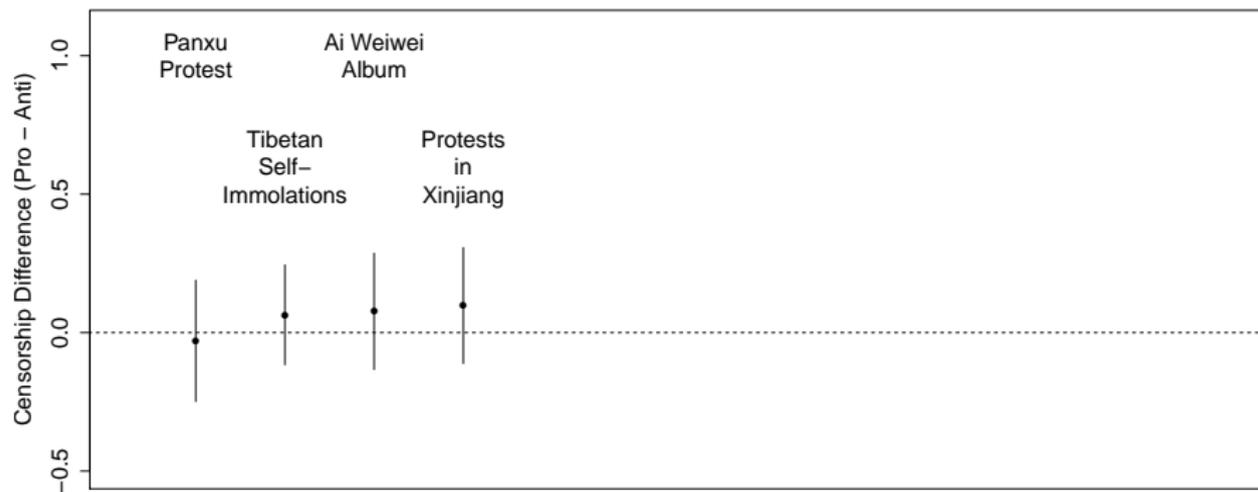
# Posts For v. Against Government: Zero Causal Effect



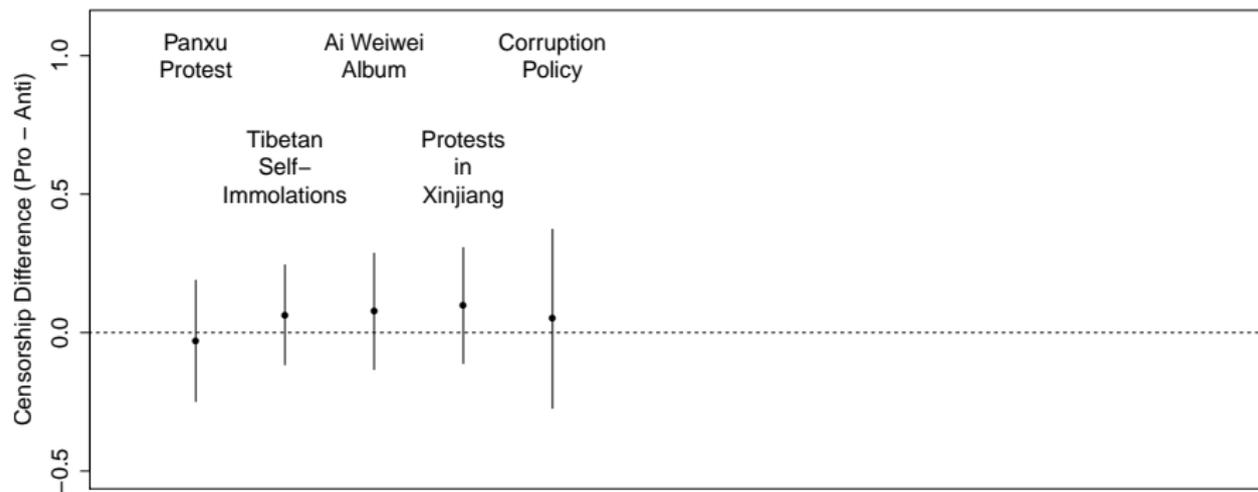
# Posts For v. Against Government: Zero Causal Effect



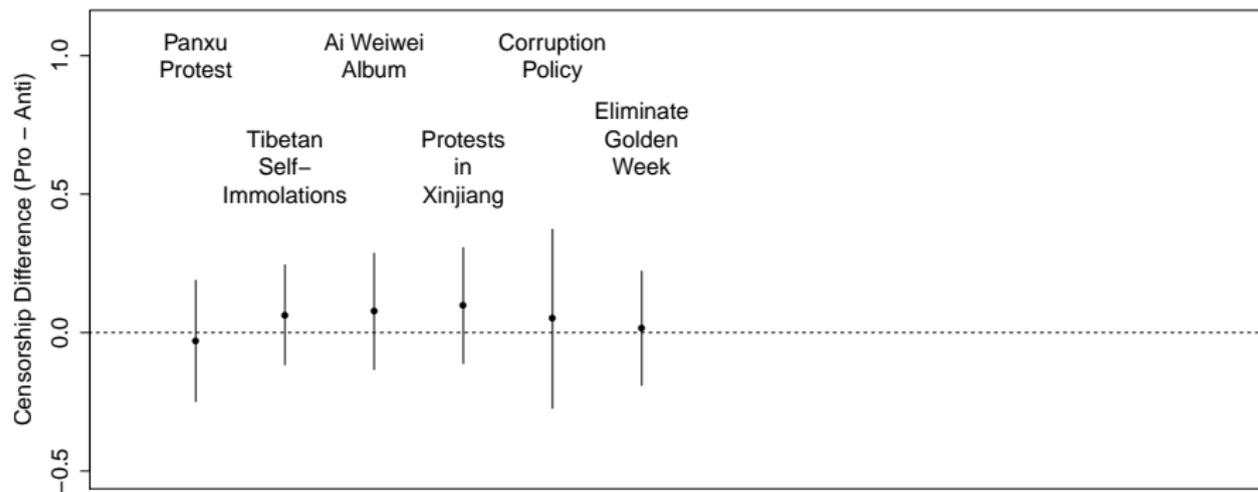
# Posts For v. Against Government: Zero Causal Effect



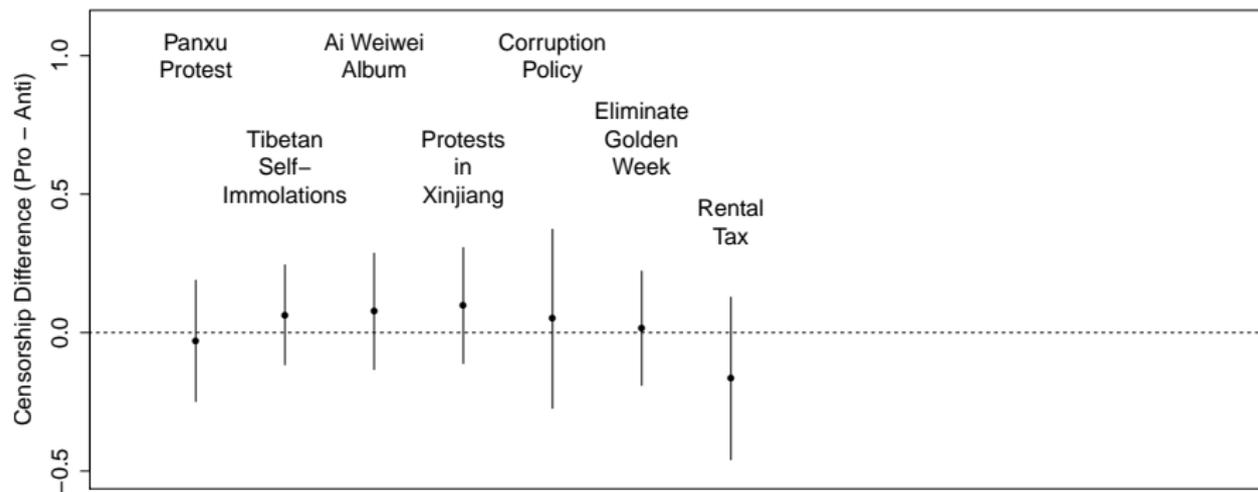
# Posts For v. Against Government: Zero Causal Effect



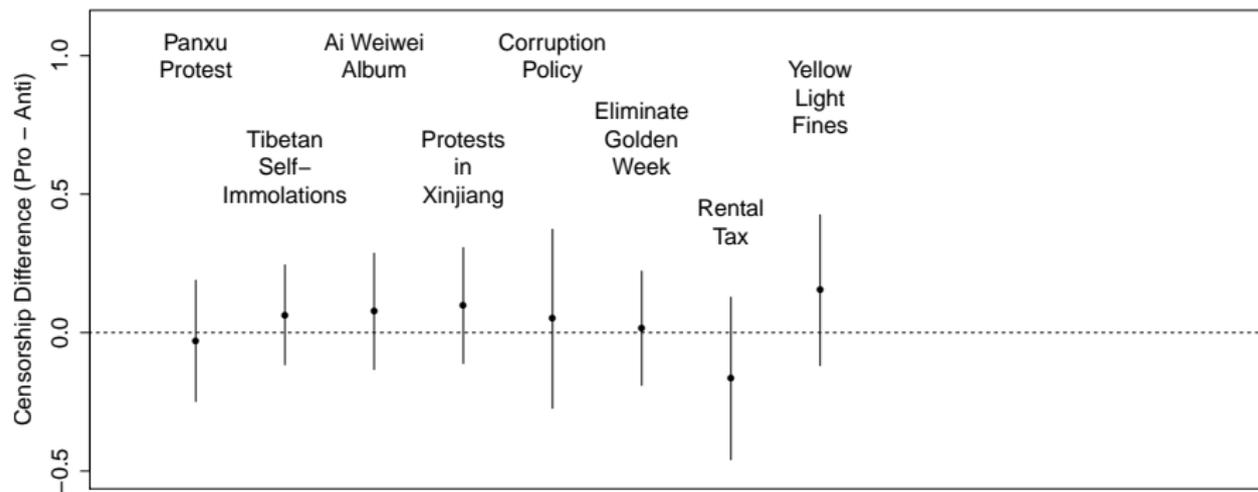
# Posts For v. Against Government: Zero Causal Effect



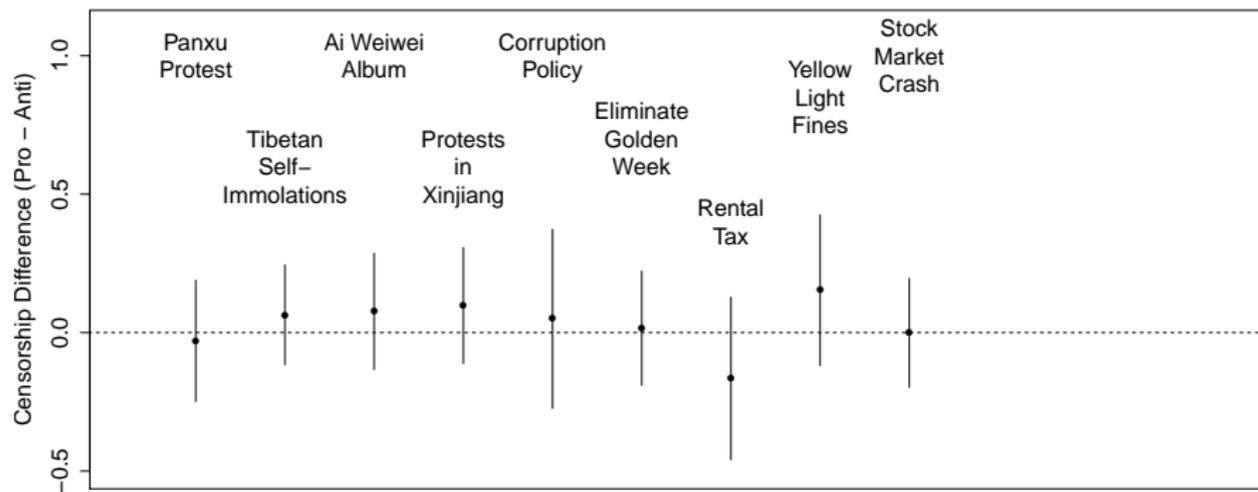
# Posts For v. Against Government: Zero Causal Effect



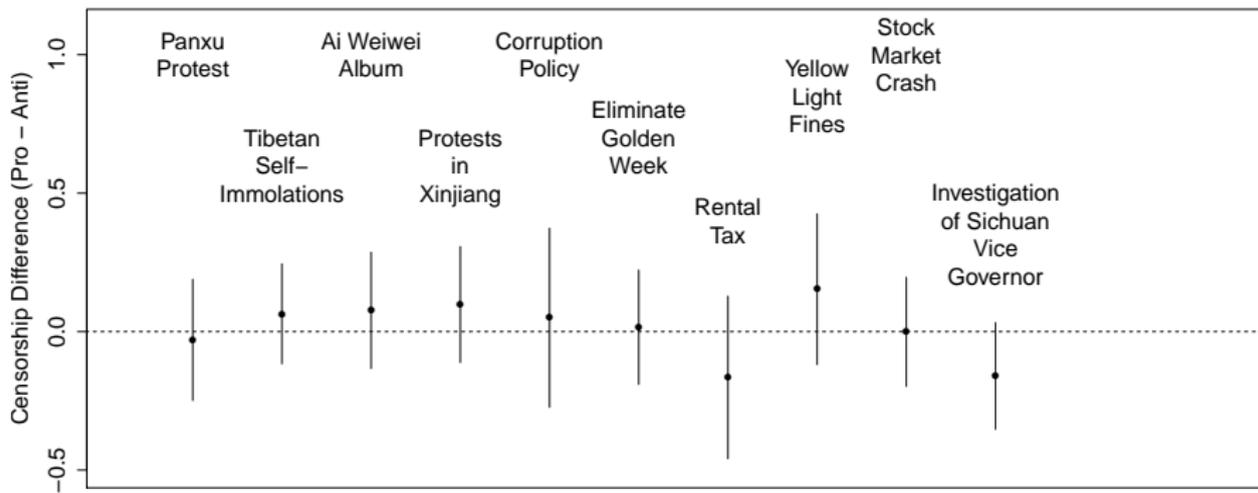
# Posts For v. Against Government: Zero Causal Effect



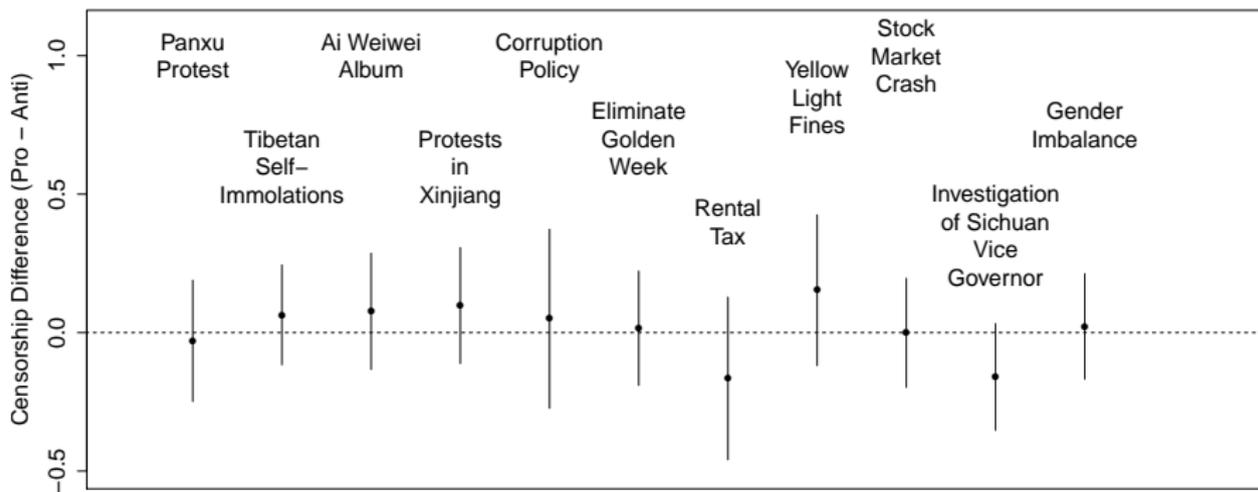
# Posts For v. Against Government: Zero Causal Effect



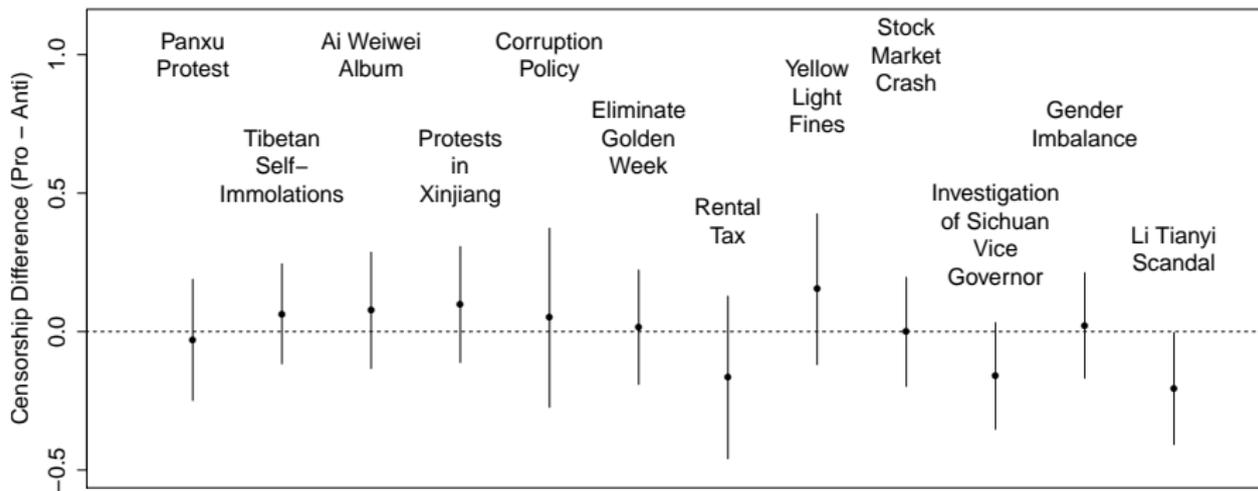
# Posts For v. Against Government: Zero Causal Effect



# Posts For v. Against Government: Zero Causal Effect

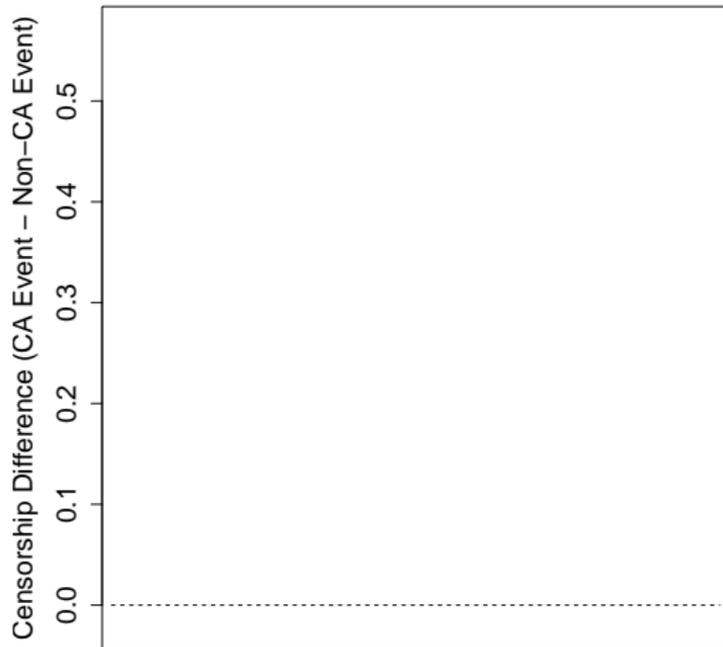


# Posts For v. Against Government: Zero Causal Effect

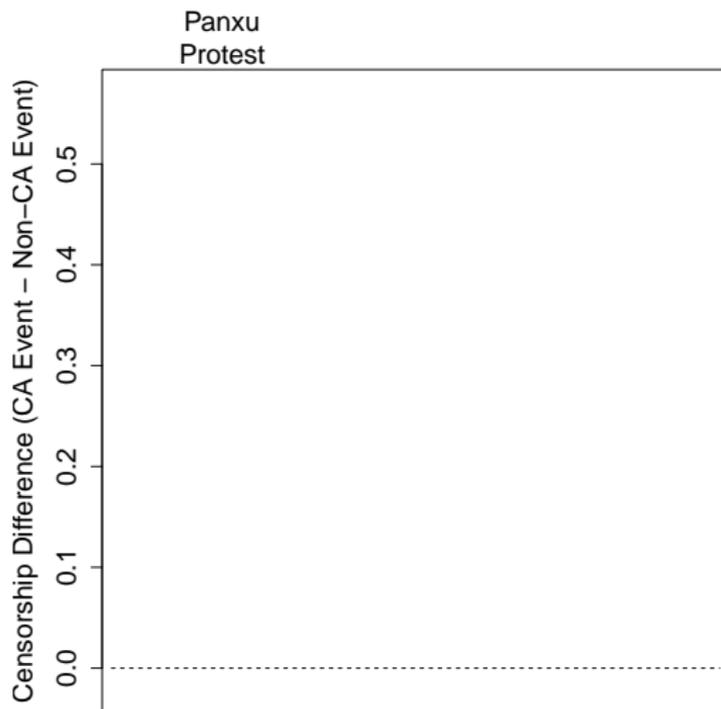


# Collective Action Events: Large Causal Effect

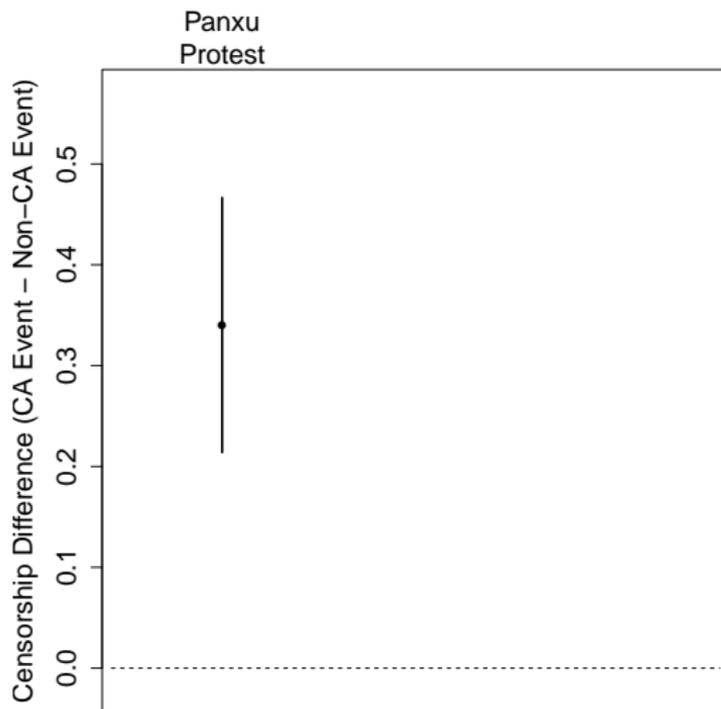
## Collective Action Events: Large Causal Effect



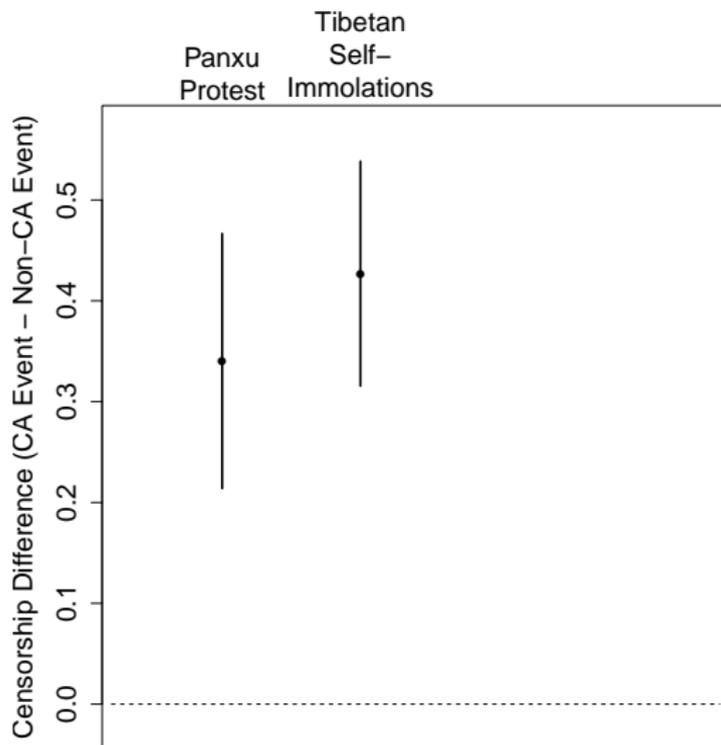
# Collective Action Events: Large Causal Effect



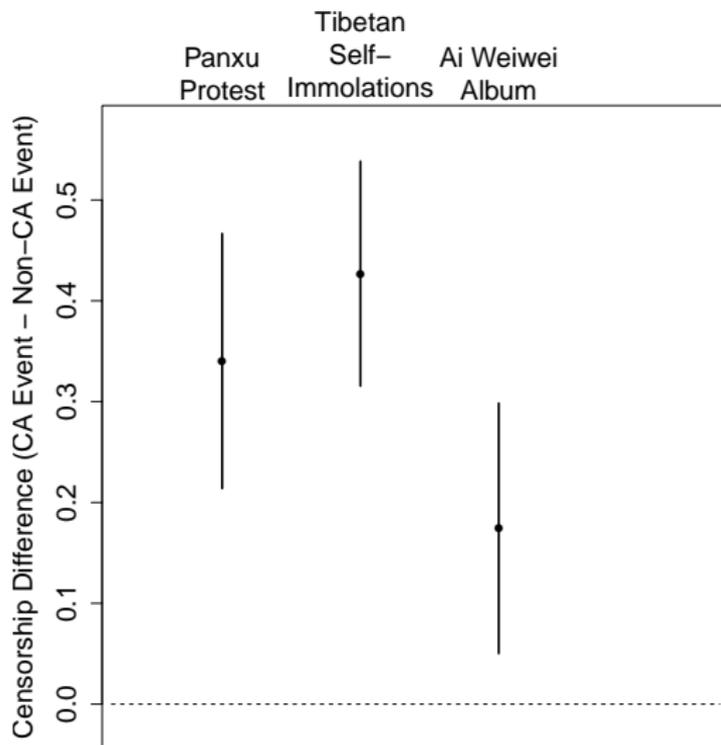
## Collective Action Events: Large Causal Effect



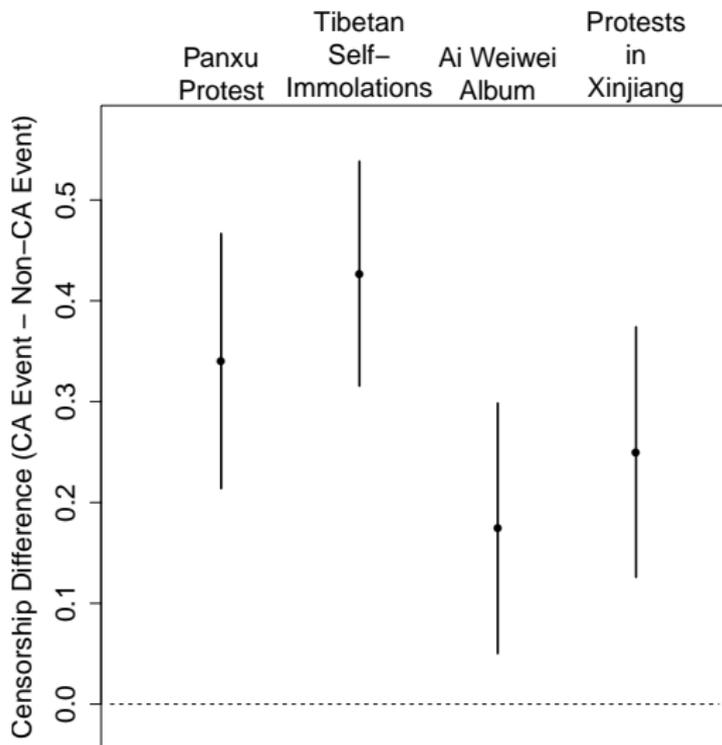
# Collective Action Events: Large Causal Effect



# Collective Action Events: Large Causal Effect



# Collective Action Events: Large Causal Effect



# Evading Automated Review

# Evading Automated Review

Example Substitution 1:

# Evading Automated Review

Example Substitution 1:

自由

# Evading Automated Review

Example Substitution 1:

自由

“Freedom”

# Evading Automated Review

Example Substitution 1:

自由

“Freedom”

**GENSORED**

# Evading Automated Review

Example Substitution 1:

自由  
自由

“Freedom”

**GENSORED**

# Evading Automated Review

Example Substitution 1:

自由  
目田

“Freedom”

“Eye field”

**GENSORED**

# Evading Automated Review

## Example Substitution 1:

自由  
目田

“Freedom”

**GENSORED**

“Eye field” (nonsensical)

# Evading Automated Review

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**GENSORED**

“Eye field” (nonsensical)

# Evading Automated Review

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**GENSORED**

“Eye field” (nonsensical)

# Evading Automated Review

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**GENSORED**

“Eye field” (nonsensical)

## Example Substitution 2:

# Evading Automated Review

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**GENSORED**

“Eye field” (nonsensical)

## Example Substitution 2:

和谐

## Evading Automated Review

### Example Substitution 1: Homograph

自由  
目田

“Freedom”

**GENSORED**

“Eye field” (nonsensical)

### Example Substitution 2:

和谐

“Harmonious [Society]” (official slogan)

# Evading Automated Review

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2:

和谐

“Harmonious [Society]” (official slogan)

**CENSORED**

# Evading Automated Review

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2:

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

# Evading Automated Review

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2:

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab”

# Evading Automated Review

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**GENSORED**

“Eye field” (nonsensical)

## Example Substitution 2:

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**GENSORED**

“River crab” (irrelevant)

## Evading Automated Review

### Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

### Example Substitution 2: Homophone (both sound like “hexie”)

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab” (irrelevant)

## Evading Automated Review

### Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

### Example Substitution 2: Homophone (both sound like “hexie”)

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab” (irrelevant)

They can't follow the conversation; our new keyword methods can!

# Evading Automated Review

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2: Homophone (both sound like “hexie”)

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab” (irrelevant)

They can't follow the conversation; our new keyword methods can!

The same task:

## Evading Automated Review

### Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

### Example Substitution 2: Homophone (both sound like “hexie”)

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab” (irrelevant)

They can't follow the conversation; our new keyword methods can!

The same task: (1) the analyst's job,

# Evading Automated Review

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2: Homophone (both sound like “hexie”)

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab” (irrelevant)

They can't follow the conversation; our new keyword methods can!

The same task: (1) the analyst's job, (2) language drift  
(#BostonBombings  $\rightsquigarrow$  #BostonStrong),

# Evading Automated Review

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2: Homophone (both sound like “hexie”)

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab” (irrelevant)

They can't follow the conversation; our new keyword methods can!

The same task: (1) the analyst's job, (2) language drift  
(#BostonBombings ~> #BostonStrong), (3) People hiding in plain sight,

# Evading Automated Review

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2: Homophone (both sound like “hexie”)

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab” (irrelevant)

They can't follow the conversation; our new keyword methods can!

The same task: (1) the analyst's job, (2) language drift

(#BostonBombings  $\rightsquigarrow$  #BostonStrong), (3) People hiding in plain sight,

(4) Look-alike modeling,

# Evading Automated Review

## Example Substitution 1: Homograph

自由  
目田

“Freedom”

**CENSORED**

“Eye field” (nonsensical)

## Example Substitution 2: Homophone (both sound like “hexie”)

和谐  
河蟹

“Harmonious [Society]” (official slogan)

**CENSORED**

“River crab” (irrelevant)

## They can't follow the conversation; our new keyword methods can!

The same task: (1) the analyst's job, (2) language drift  
(#BostonBombings  $\rightsquigarrow$  #BostonStrong), (3) People hiding in plain sight,  
(4) Look-alike modeling, (5) Starting point for sophisticated automated  
text analysis

# Conclusions

# Conclusions

- The Goals of Chinese Censorship:

# Conclusions

- The Goals of Chinese Censorship:
  - For the Chinese people: Individually free, Collectively in Chains

# Conclusions

- The Goals of Chinese Censorship:
  - For the Chinese people: Individually free, Collectively in Chains
  - Unambiguous evidence from: observational study, participant observation, & randomized experiments

# Conclusions

- The Goals of Chinese Censorship:
  - For the Chinese people: Individually free, Collectively in Chains
  - Unambiguous evidence from: observational study, participant observation, & randomized experiments
  - Makes Social Media Actionable for us:

# Conclusions

- The Goals of Chinese Censorship:
  - For the Chinese people: Individually free, Collectively in Chains
  - Unambiguous evidence from: observational study, participant observation, & randomized experiments
  - Makes Social Media Actionable for us:
    - Enables detailed, continuous, predictive monitoring of every Chinese government, leader, policy, dissident, scandal, company, etc.

# Conclusions

- The Goals of Chinese Censorship:
  - For the Chinese people: Individually free, Collectively in Chains
  - Unambiguous evidence from: observational study, participant observation, & randomized experiments
  - Makes Social Media Actionable for us:
    - Enables detailed, continuous, predictive monitoring of every Chinese government, leader, policy, dissident, scandal, company, etc.
    - Predicts state actions when other sources are silent

# Conclusions

- The Goals of Chinese Censorship:
  - For the Chinese people: Individually free, Collectively in Chains
  - Unambiguous evidence from: observational study, participant observation, & randomized experiments
  - Makes Social Media Actionable for us:
    - Enables detailed, continuous, predictive monitoring of every Chinese government, leader, policy, dissident, scandal, company, etc.
    - Predicts state actions when other sources are silent
    - Applicable to some other countries

# Conclusions

- The Goals of Chinese Censorship:
  - For the Chinese people: Individually free, Collectively in Chains
  - Unambiguous evidence from: observational study, participant observation, & randomized experiments
  - Makes Social Media Actionable for us:
    - Enables detailed, continuous, predictive monitoring of every Chinese government, leader, policy, dissident, scandal, company, etc.
    - Predicts state actions when other sources are silent
    - Applicable to some other countries
- New methods developed for:

# Conclusions

- The Goals of Chinese Censorship:
  - For the Chinese people: Individually free, Collectively in Chains
  - Unambiguous evidence from: observational study, participant observation, & randomized experiments
  - Makes Social Media Actionable for us:
    - Enables detailed, continuous, predictive monitoring of every Chinese government, leader, policy, dissident, scandal, company, etc.
    - Predicts state actions when other sources are silent
    - Applicable to some other countries
- New methods developed for:
  - Following conversations when people try to evade censors

# Conclusions

- **The Goals of Chinese Censorship:**
  - For the Chinese people: Individually free, Collectively in Chains
  - Unambiguous evidence from: observational study, participant observation, & randomized experiments
  - **Makes Social Media Actionable for us:**
    - Enables detailed, continuous, predictive monitoring of every Chinese government, leader, policy, dissident, scandal, company, etc.
    - Predicts state actions when other sources are silent
    - Applicable to some other countries
- **New methods developed for:**
  - Following conversations when people try to evade censors
  - Automated keyword recommender for government analysts

# Conclusions

- **The Goals of Chinese Censorship:**
  - For the Chinese people: Individually free, Collectively in Chains
  - Unambiguous evidence from: observational study, participant observation, & randomized experiments
  - **Makes Social Media Actionable for us:**
    - Enables detailed, continuous, predictive monitoring of every Chinese government, leader, policy, dissident, scandal, company, etc.
    - Predicts state actions when other sources are silent
    - Applicable to some other countries
- **New methods developed for:**
  - Following conversations when people try to evade censors
  - Automated keyword recommender for government analysts
  - Predicting action on the ground from censorship

# Conclusions

- **The Goals of Chinese Censorship:**
  - For the Chinese people: Individually free, Collectively in Chains
  - Unambiguous evidence from: observational study, participant observation, & randomized experiments
  - **Makes Social Media Actionable for us:**
    - Enables detailed, continuous, predictive monitoring of every Chinese government, leader, policy, dissident, scandal, company, etc.
    - Predicts state actions when other sources are silent
    - Applicable to some other countries
- **New methods developed for:**
  - Following conversations when people try to evade censors
  - Automated keyword recommender for government analysts
  - Predicting action on the ground from censorship
  - Understanding large volumes of unstructured text

## For more information



[GaryKing.org](http://GaryKing.org)



[j.mp/JenPan](https://j.mp/JenPan)

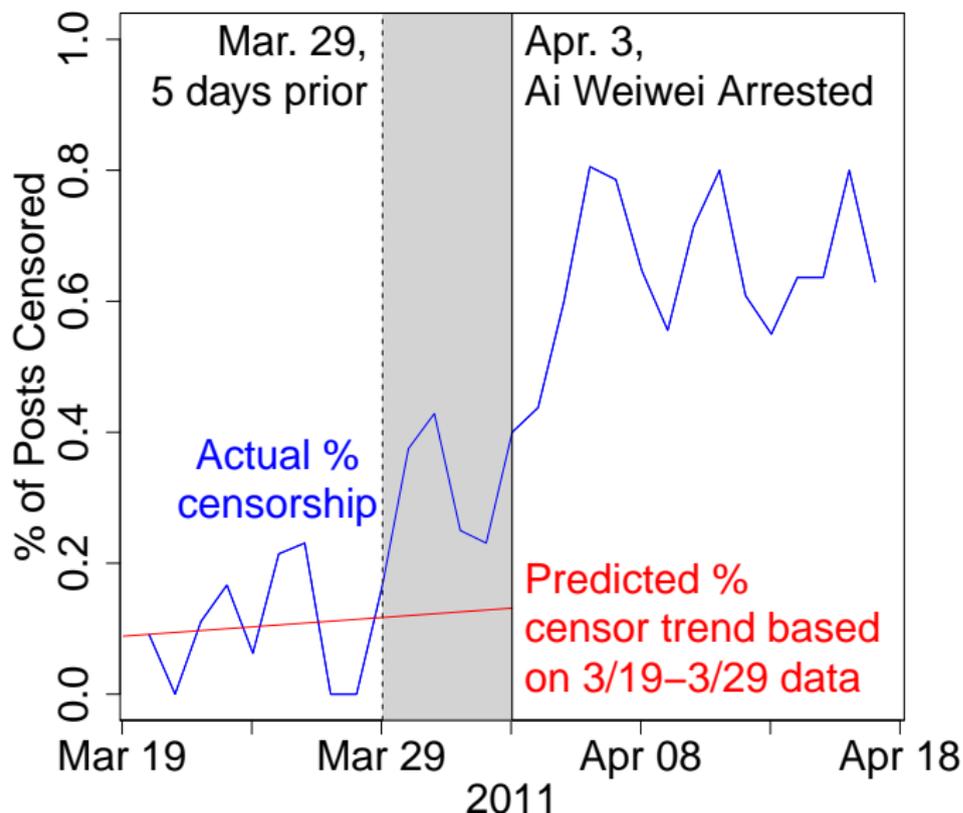


[j.mp/MollyRoberts](https://j.mp/MollyRoberts)

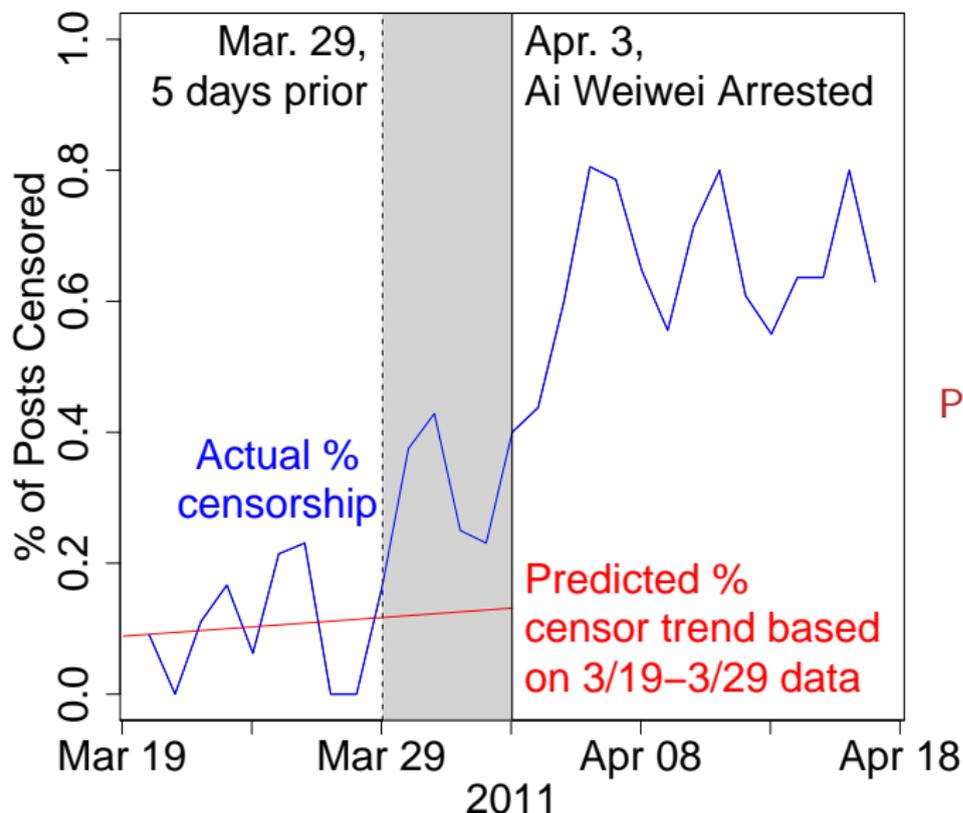
# Appendix

# Censorship to Preempt Collective Action: Ai Weiwei's Arrest

# Censorship to Preempt Collective Action: Ai Weiwei's Arrest

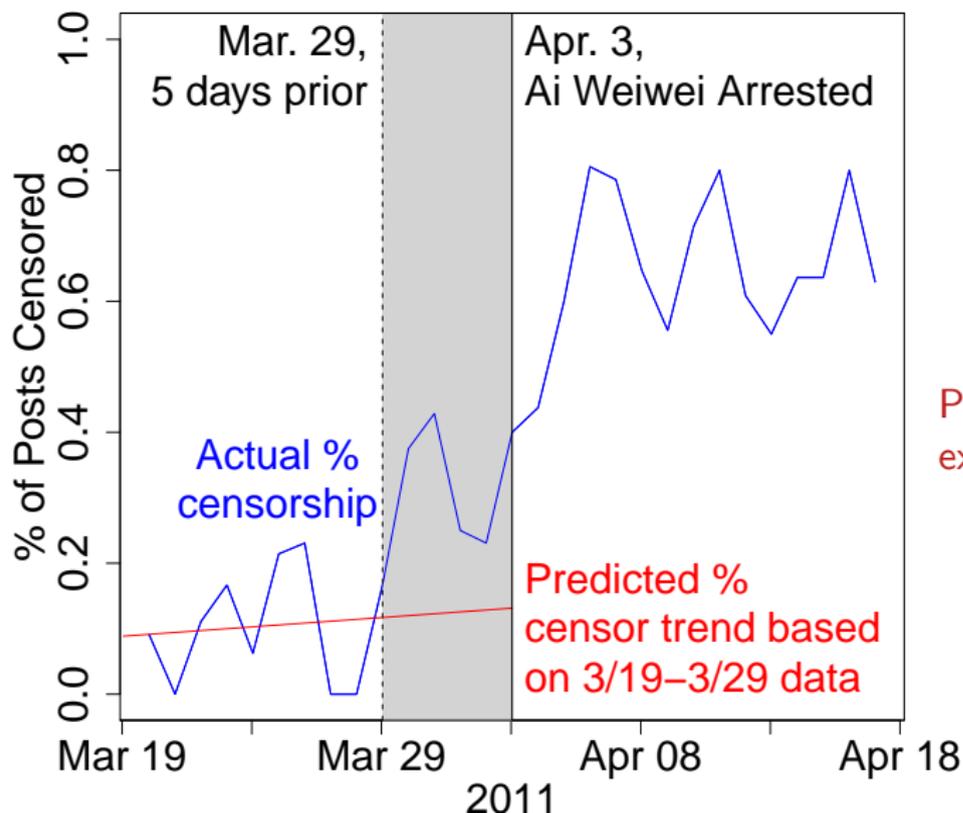


# Censorship to Preempt Collective Action: Ai Weiwei's Arrest



Placebo Test:

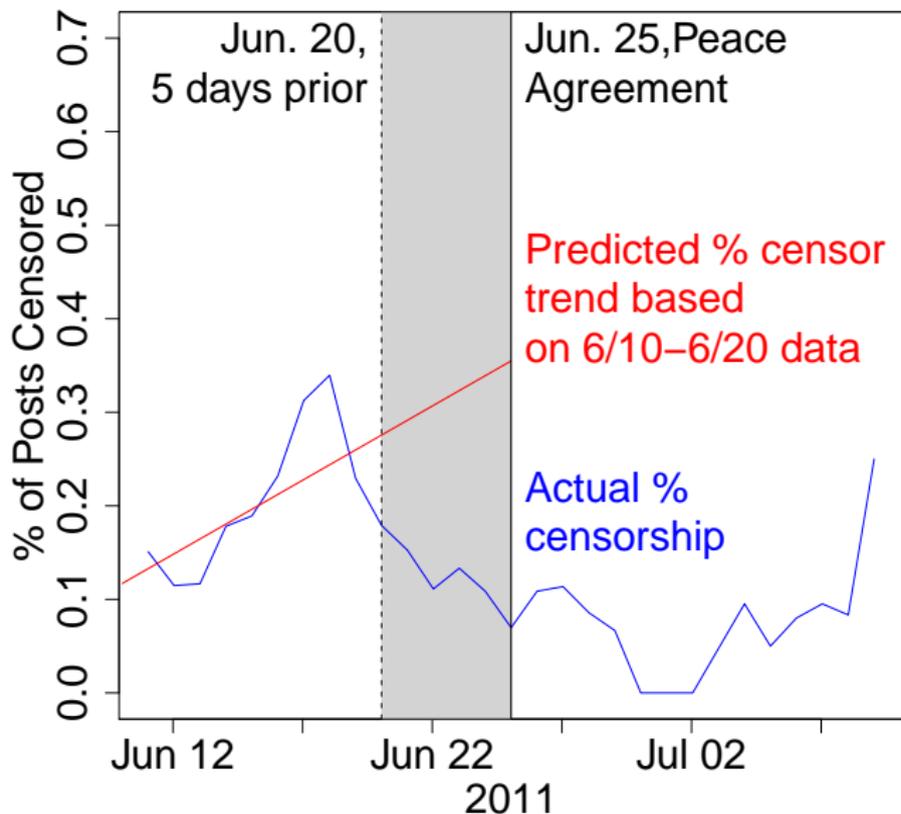
# Censorship to Preempt Collective Action: Ai Weiwei's Arrest



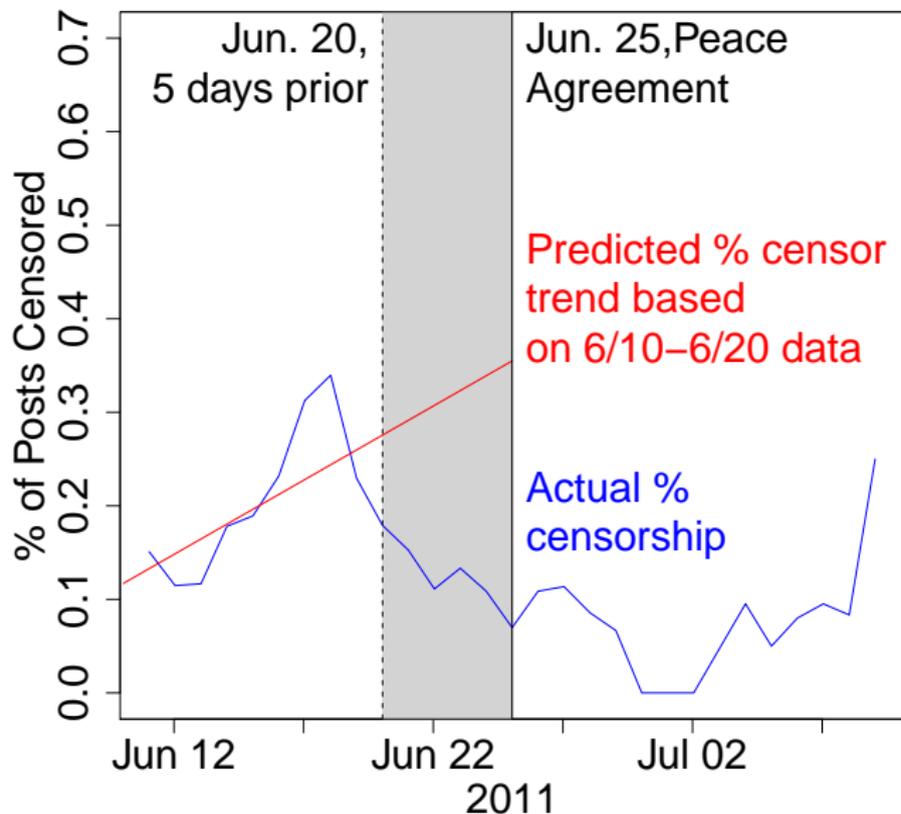
Placebo Test: Most extreme of all effects

# Predicting the South China Sea Peace Agreement

# Predicting the South China Sea Peace Agreement

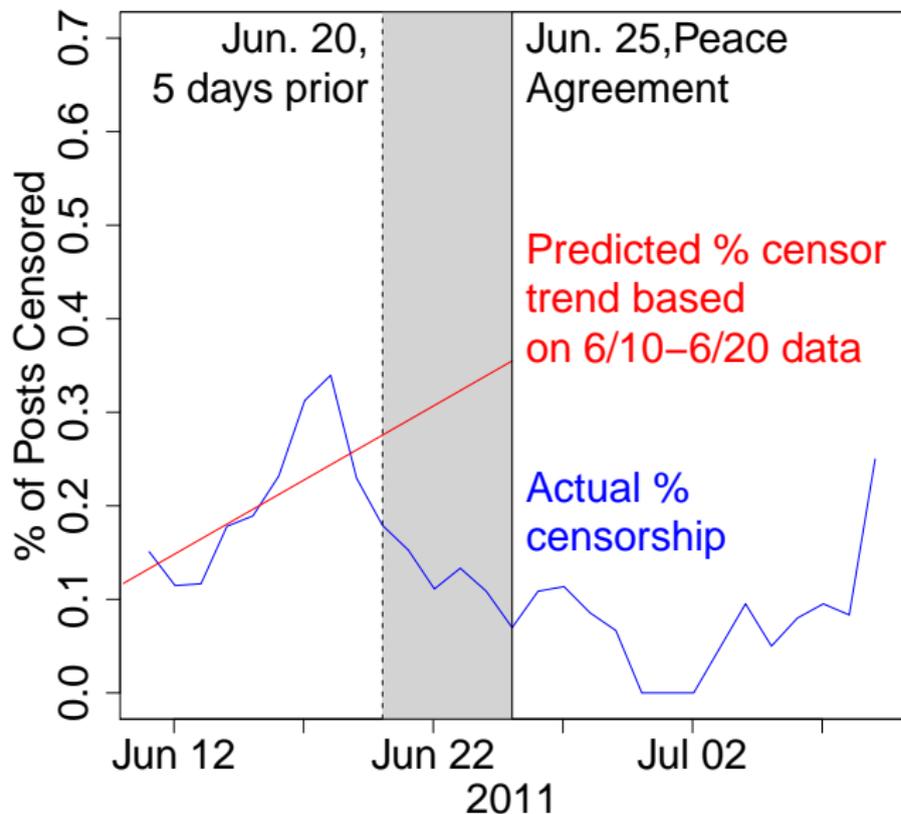


# Predicting the South China Sea Peace Agreement



Placebo Test:

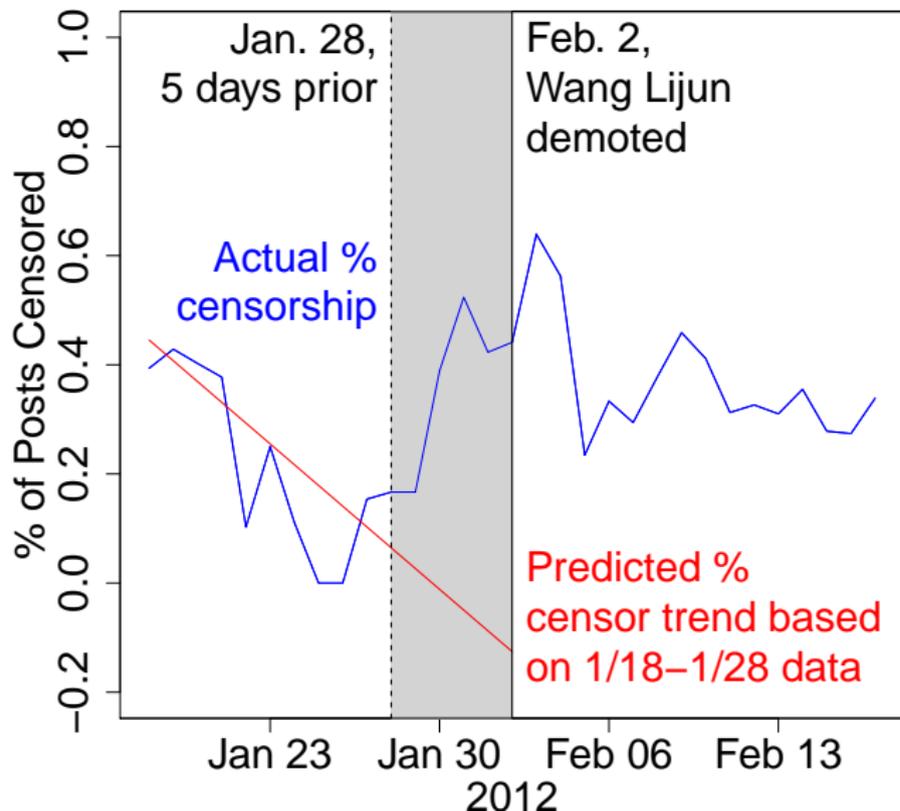
# Predicting the South China Sea Peace Agreement



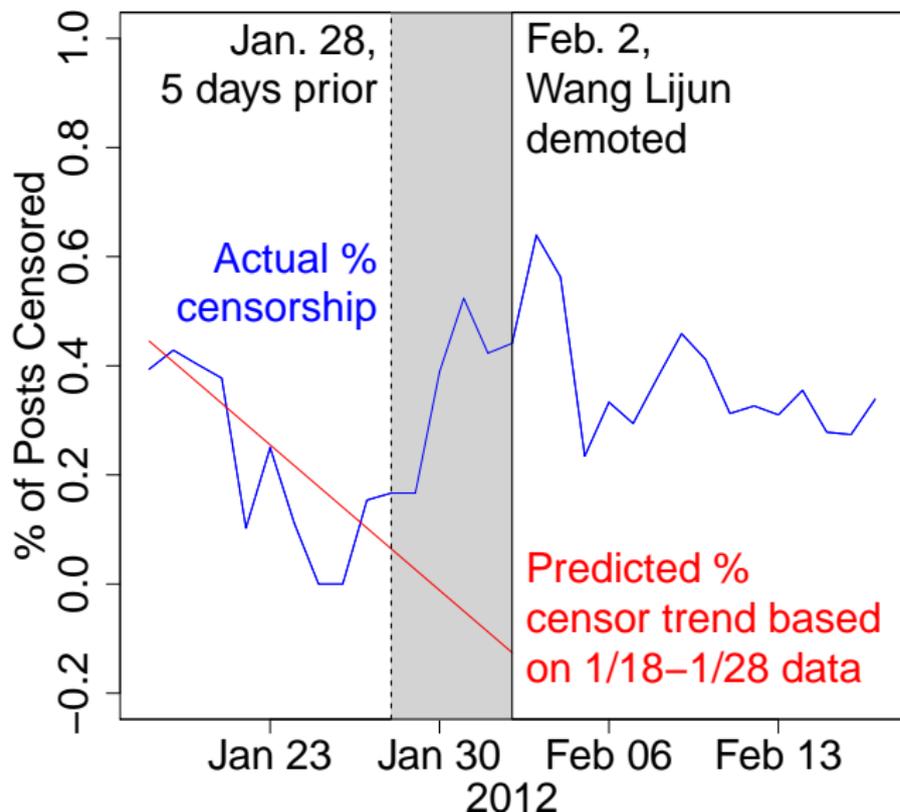
Placebo Test: Most extreme of all effects

# Censorship Pre-empting Collective Action: Wang Lijun's Demotion

# Censorship Pre-empting Collective Action: Wang Lijun's Demotion

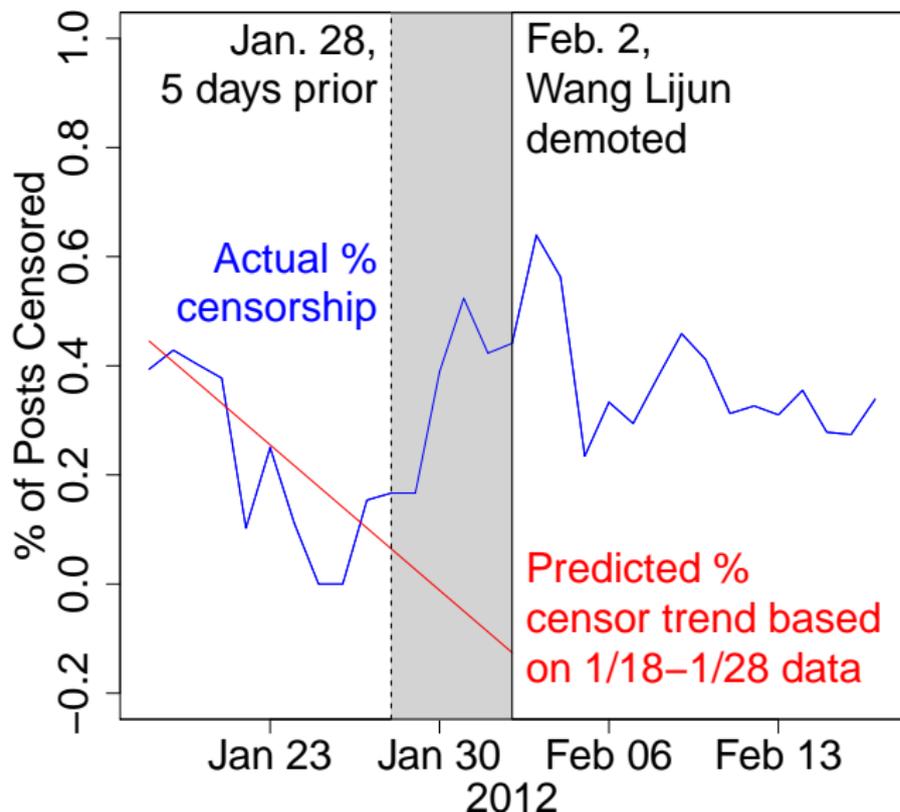


# Censorship Pre-empting Collective Action: Wang Lijun's Demotion



Placebo Test:

# Censorship Pre-empting Collective Action: Wang Lijun's Demotion



Placebo Test: Most extreme of all effects

# Uncensored Posts (w/o Collective Action Potential) Critical of the State

# Uncensored Posts (w/o Collective Action Potential) Critical of the State

		Collective Action	
		yes	no
State Critique	yes	<b>CENSORED</b>	<b>Free</b>
	no	<b>CENSORED</b>	<b>Free</b>

## Uncensored Posts (w/o Collective Action Potential) Critical of the State

This is a city government [Yulin City, Shaanxi] that treats life with contempt, this is government officials run amuck, a city government without justice, a city government that delights in that which is vulgar, a place where officials all have mistresses, a city government that is shameless with greed, a government that trades dignity for power, a government without humanity, a government that has no limits on immorality, a government that goes back on its word, a government that treats kindness with ingratitude, a government that cares nothing for posterity. . .

这是一个漠视生命的市政府[陕西省榆林市]、一个官员横行的市政府、一个没有公正的市政府，一个低级趣味的市政府，一个包二奶的市政府，一个为钱不要脸的市政府，一个为个权不要人格的市政府，一个没有血性的市政府，一个没有道德低线的市政府，一个出尔反尔的市政府，一个忘恩负义的市政府，一个不要子孙后代的市政府，一个什么怪事都出的市政府，一个什么的市政府，只要你想到的就有…

# Censored Post (with Collective Action Potential) Supporting the State

# Censored Post (with Collective Action Potential) Supporting the State

		Collective Action	
		yes	no
State Critique	yes		<b>Free</b>
	no		<b>Free</b>

## Censored Post (with Collective Action Potential) Supporting the State

The bombing led not only to the tragedy of his death but the death of many government workers. Even if we can verify what Qian Mingqi said on Weibo that the building demolition caused a great deal of personal damage, we should still condemn his extreme act of retribution. . . . The government has continually put forth measures and laws to protect the interests of citizens in building demolition. And the media has called attention to the plight of those experiencing housing demolition. The rate at which compensation for housing demolition has increased exceeds inflation. In many places, this compensation can change the fate of an entire family.

爆炸案造成他本人和多名政府工作人员死伤的悲剧，即使钱明奇在微博里所称拆迁造成的个人损失是属实的，我们也应谴责他的极端报复行为. . . . 政府在连续出台保护被拆迁者利益的政策法规，媒体也在为公平对待被拆迁者大声疾呼，各地拆迁补偿款的上升速度，大多高于商品房售价上升的速度，在不少地方，补偿款已经足以改变一个家庭的命运。