


RESEARCH ARTICLE

Selection and Description Bias in Protest Reporting by Government and News Media on Weibo

Han Zhang¹ , Yao Lu² and Rui Bai³

¹Division of Social Science, Hong Kong University of Science and Technology, Hong Kong, China, ²Department of Sociology, Columbia University, New York, USA and ³Data Science Institute, Columbia University, New York, USA

Corresponding author: Han Zhang, email: zhangh@ust.hk

Abstract

Extensive research in Western societies has demonstrated that media reports of protests have succumbed to selection and description biases, but such tendencies have not yet been tested in the Chinese context. This article investigates the Chinese government and news media's selection and description bias in domestic protest events reporting. Using a large protest event data set from Weibo (CASM-China), we found that government accounts on Weibo covered only 0.4 per cent of protests while news media accounts covered 6.3 per cent of them. In selecting events for coverage, the news media accounts tacitly struck a balance between newsworthiness and political sensitivity; this led them to gravitate towards protests by underprivileged social groups and shy away from protests targeting the government. Government accounts on Weibo, on the other hand, eschewed reporting on violent protests and those organized by the urban middle class and veterans. In reporting selected protest events, both government and news media accounts tended to depoliticize protest events and to frame them in a more positive tone. This description bias was more pronounced for the government than the news media accounts. The government coverage of protest events also had a more thematic (as opposed to episodic) orientation than the news media.

摘要

基于西方社会的研究发现媒体对抗议的报道有选择和描述偏差，但是这些研究发现尚未在中国语境中得以验证。这篇文章探讨中国政府和媒体在报道抗议事件时的选择和描述偏差。利用一个大规模的基于新浪微博的抗议数据库 (CASM-China)，本文发现微博上的政府账户仅报道了 0.4% 的抗议事件，而新闻媒体的微博账户报道了 6.3% 的抗议事件。新闻媒体选择报道事件时平衡新闻性和政治敏感性；它们倾向于报道社会底层人群的抗议，但是不去报道针对政府的抗议。微博上的政府媒体倾向于不报道暴力事件，以及中产阶级和退伍军人的抗议。关于描述偏差，本文发现政府和新闻媒体账户均采用去政治化的报道方式，同时使用更正面的语气。政府账户的报道偏差大于新闻媒体的报道偏差。政府账户同时倾向于使用更一般性（而非特殊性）的语言来描述抗议

Keywords: protest; news media; social media; media bias; Weibo; machine learning; text analysis

关键词: 抗议; 新闻媒体; 社交媒体; 媒体偏差; 微博; 机器学习; 文本分析

The media is crucial to the development and outcomes of social protests.¹ The general public becomes aware of protests primarily through the media, which has a broad reach that transcends individuals physically present at the protests. Scholarship on social movements has traditionally identified the media as the most important channel shaping people's participation in protests and their legitimacy, thereby influencing the public support of protests.² This channel also mediates the relationship between state and societal actors (protesters): autocratic states seek to contain

1 Caren, Andrews and Lu 2020.

2 Oliver and Maney 2000.

movement diffusion and maintain social stability by prohibiting the media from reporting on contentious events,³ whereas democratic states attempt to influence the ways in which the media reports protest events. Koopmans goes so far as to claim that the media has become the major battleground between protesters and the state.⁴

However, the burgeoning literature on the protest–media relationship demonstrates that media reports of protests are neither representative nor neutral. Two main forms of bias exist in the media’s representation of protests, namely selection bias and description bias.⁵ These biases can be described as the media covering only a small, selected number of protests (selection bias) and, for the events selected, portraying events differently from the way protest participants experience them (description bias). Research on media selection and description bias has focused on democratic societies and highlighted the importance of factors such as news value and the news cycle in media reporting and its content.⁶ In an authoritarian regime such as China, however, a unique set of factors may shape media selection and description bias. The specific circumstances of authoritarian settings suggest that media bias may stem largely from an understanding of the government’s restrictions and priorities, rather than from newsworthiness and editors’ discretion.⁷

Does the media in China selectively report protests? Do media reports differ from the individuals’ portrayal of the events? Answers to these questions can improve our understanding of the interactions between the state, the media and society in China. However, little is known about this topic due to the limited availability of data on protest events in the country. To the best of the authors’ knowledge, this study is among the first to systematically investigate these questions.

Addressing these questions in China requires a modified methodological approach that is informed by the existing literature on Western societies. In the West, scholars often collect protest event data from newspapers. But this is not viable in autocratic societies because the number of protest events appearing in the newspapers is extremely limited relative to the actual number of protests on the ground. Print media is heavily regulated in China and is largely prohibited from reporting on collective action. For instance, official statistics from the Chinese Ministry of Public Security reported 87,000 “mass incidents” in 2005; but fewer than 500 protests per year have appeared in newspapers per year across the entire country.⁸

In contrast, social media has grown substantially in the past few decades, thanks to the advent of modern communication and digital technologies. While still under state control, social media is not regulated to the same rigor as print media and has emerged as a prominent platform for information dissemination and mobilization.⁹ Social media has reinvented reporting by allowing protesters to directly publicize their efforts and mobilize offline action.¹⁰ This phenomenon takes place even in China.¹¹ Moreover, social media has a broader reach as it offers a more widely available, decentralized channel of information flow than traditional mass media outlets. As such, even traditional mass media outlets (e.g. newspapers) have become active players in the social media world by digitally disseminating their content. Local governments in China have also moved towards “consultative authoritarianism” on social media by collecting individual grievances and posting resolution plans.¹²

3 Hale 2013.

4 Koopmans 2004.

5 McCarthy, McPhail and Smith 1996; Oliver and Maney 2000.

6 Andrews and Caren 2010.

7 Ong 2015.

8 Shao 2017; Chen, Chih-Jou Jay 2020.

9 Stockmann and Luo 2017; Roberts 2018.

10 Earl 2015.

11 Qin, Strömberg and Wu 2018.

12 Distelhorst and Hou 2017.

This article uses CASM-China, a unique publicly available data set that contains information on more than 136,330 offline protest events in China from 2010 to mid-2017, with detailed geographical information down to the county level.¹³ The data set uses a two-step deep-learning algorithm that identifies offline protest events from 9.5 million Weibo 微博 (a Twitter analogue in China) posts that contain protest-related words. Zhang and Pan have performed extensive validations to demonstrate that such machine identification of protests is reasonably accurate and not subject to major biases brought about by censorship. In this article, for each event in CASM-China posted by individuals (i.e. non-government and non-media accounts, mostly protesters themselves but occasionally bystanders who witnessed the protests), we first examined whether the event was reported by government or news media accounts on Weibo and then considered the factors predicting the probability of reporting (selection bias). Second, for events that were reported, we further studied whether and how government and news media accounts portrayed the protests as compared to individuals (description bias). We distinguished between news media and government accounts in all analyses and, in additional analysis, we further differentiated different types of news media on Weibo, based on their embeddedness with the state.

In general, the results demonstrate notable selection and description bias in media reports of protest events on Chinese Weibo. Selection bias for news media and government accounts have similarities, but also differ in certain ways. Description bias, on the other hand, is more pronounced in governmental than news media reporting. Among news media, those closer to the state (being the government's mouthpiece) exhibit selection and description biases similar to those of government accounts, whereas news media established by persons unaffiliated with the state showed the lowest level of selection and description biases, similar to the original descriptions by individuals who participated or witnessed the protests.

Literature

Limited media attention on protests in China and why it still matters

Social protests are widespread in China. Reports published by China's Ministry of Public Security point to 87,000 "mass incidents" occurring in 2005. Although the government ceased publishing aggregate counts of mass incidents after 2005, a widely cited number in the literature points to 180,000 mass incidents in 2010.¹⁴ The frequency of these protests has prompted the Chinese government to make a reduction in the number of protests, or "maintaining stability" (*weiwen* 维稳), its top political priority.¹⁵

Nevertheless, ordinary citizens who are unfamiliar with the issues and rely primarily on traditional mass media may not believe that China faces a large number of protests. In the two most comprehensive data collection efforts on protest events reported in newspapers, Shao identified 5,708 events between 1998 and 2014,¹⁶ and Chen collected data on more than 10,000 events from 2000 to 2018.¹⁷ On average, these two studies have identified only 400 to 500 protests each year from more than 700 newspapers in mainland China,¹⁸ which translates to less than one event per year per newspaper. Ong¹⁹ and Elfstrom and Kuruvilla²⁰ collected protest events from

13 Zhang and Pan 2019.

14 Lee and Zhang 2013; Goebel and Ong 2012.

15 Wang and Minzner 2015.

16 Shao 2017.

17 Chen, Chih-Jou Jay 2020.

18 Both studies used WiseNews, which is the largest electronic searchable database in China and contains over 870 newspapers.

19 Ong 2015.

20 Elfstrom and Kuruvilla 2014.

mixed sources, including internet websites and mass media, but their data sets also contain information on only several hundred protests per year.

Despite the strikingly low coverage of protests, news media reports of these events still matter. Paradoxically, the general lack of attention paid to protests by traditional mass media means that, when protests do make their way into newspapers or television, they tend to garner tremendous attention and public support. Upper-level government is compelled to intervene, often by instructing local governments to make concessions to protesters. Cai has found that media exposure is among the most important factors facilitating the success of protests in China.²¹ Its effect on protest success is on par with factors such as the number of casualties and the size of the protest.²²

Media reports on protests also matter in that they shape public understanding of existing social problems and governments' limits and disseminate protest-related information (issues, tactics, etc.). Importantly, such media coverage signals political agendas and openings, legitimacy, and state tolerance of protests at the national and local level. Such processes may inspire fellow citizens with similar grievances to mobilize. In extreme cases, media reporting on protests may facilitate waves of protests, as in the widely cited metaphor "a single spark can start a prairie fire."²³

The past decade has witnessed a proliferation of social media in China and around the world. Social media platforms have provided ordinary citizens with a new channel for mobilization, and this topic has been extensively studied in different parts of the world.²⁴ Traditional mass media and governments have also emerged as active players in social media spaces. Nearly all traditional mass media outlets in China, such as newspapers and television, have set up and maintained their own social media accounts; such a strategy has proven useful for "saving the newspaper."²⁵ In 2012, only three of the twenty news media accounts with the most followers on Weibo were owned by the state. Four years later, the twenty news media accounts with the most followers all represented traditional news media organizations that had existed long before Weibo and were all affiliated with the state.²⁶ Local governments and their affiliated departments and agencies (e.g. environmental departments or courts) entered the social media world relatively late, but social media has quickly become the preferred channel for direct government-to-public communication.²⁷ The main functions of government accounts on Weibo have been to create positive propaganda messages for the party-state, interpret central policies, and influence public opinion, especially during unexpected events.²⁸

Given the aforementioned circumstances in China, this article focuses on news media and government accounts on social media and examines potential selection and reporting biases – that is, whether certain protests receive greater media attention than others, and whether there is any bias in the media's portrayal of protest events.

Media biases in social movements

Media biases in social movements reporting have been extensively researched in the Western context.²⁹ This literature focuses predominantly on two types of media biases: selection bias and

21 Cai 2010.

22 On a related note, studies have shown that protests are usually more effective than institutional channels of claim-making such as *xinfang* 信访 ("Letters and Petitions"; a government agency handling citizen complaints) or lawsuits in China. Ibid.

23 Lorentzen 2017.

24 Wolfsfeld, Segev and Sheaffer 2013.

25 Hong 2012; Ju, Jeong and Chyi 2014.

26 Repnikova and Fang 2018.

27 Ibid.

28 See State Council 2018.

29 There is a larger literature on media biases in other domains, such as media biases in covering political news (Eberl, Boomgaarden and Wagner 2017) or other agendas. This study confines the literature to media biases in reporting social movements.

description bias. This research adopts the same definitions as Earl et al.³⁰ Selection bias aims to measure “which subset of events are covered.” Description bias aims to measure “the veracity of the coverage.” A thorough review of media bias in protest reporting can be found in Hutter.³¹

For the causes of media selection biases, research has focused on newsworthiness and organizational factors.³² In democratic countries, news organizations play a prominent “gatekeeper” role and make decisions about coverage based on the newsworthiness of protest events. Factors contributing to the newsworthiness of a protest include its spatial proximity to the news agency, its relationship to issue attention cycles,³³ the size of the protest and/or the size of the movements and organizations involved in it,³⁴ whether any violent or disruptive action has taken place,³⁵ the existence of counterdemonstrations,³⁶ and the protest’s connection to current and significant issues for the government or local legislatures, among others. Andrews and Caren³⁷ further show that protests headed by professional social movement organizations tend to garner more attention than those organized by confrontational volunteer-led groups.

In comparison, structural factors – “the broader structure of power relations in society” play a less pronounced and consistent role in media selection biases in the Western world, where the media is independent of the state.³⁸ However, these factors may carry a heavy weight in reporting of protests in authoritarian regimes. In these settings, the biggest structural constraint is the state’s regulation and sometimes direct influence in what and how protests should be reported. This article brings the state back in the dialogue and proposes a state-embedded view of media biases. This view leads to the prediction that the selection biases of protest reporting depend on their embeddedness to the state, with media closer to the state exhibiting a higher degree of biases and those closer to society (i.e. individual users) displaying a lower level of selection biases.

Traditional news media in democratic societies is also fraught with *description* bias when reporting on protests. The media generally portrays the “hard news” aspect of protests (the “who, what, when, where and why” of the protest) in a relatively unbiased fashion, especially when the events are organized by large, credible organizations. The “soft news” dimension, on the other hand, is subject to greater description bias. Such bias often involves omission of specific information rather than purposive misrepresentation and/or distortion of information. Another source of bias emanates from the framing of protests: media reports are sometimes framed to appeal to their audience rather than to address the true cause underlying the protests. Such bias can undermine the movements’ agenda and affect the public’s interpretation of the movement.³⁹

Similarly, we also posit that description biases will also likely differ depending on the actor’s embeddedness to the state. We believe that such a state-embedded view makes a contribution to the literature by providing a theory that can be extended to other autocratic regimes marked by tight media regulations. The next sections offer a typology and lay out our state-embedded theory of media biases.

30 Earl et al. 2004.

31 Hutter 2014.

32 Smith et al. 2001.

33 McCarthy, McPhail and Smith 1996.

34 Hug and Wisler 1998.

35 Barranco and Wisler 1999.

36 Oliver and Maney 2000.

37 Andrews and Caren 2010.

38 Smith et al. 2001, 1403.

39 Ibid.

Typology of actors on Chinese social media

On Chinese Weibo, many accounts are owned by the government or CCP agencies, such as local governments, people's congresses, the courts, police departments, and environmental protection departments. These government agencies and their social media accounts do not act in the same way as traditional media actors. Rather, they resort to social media as a tool to boast its policy effectiveness and promote its public image. News media accounts for an even larger share of Weibo accounts. Chinese news media operates under distinct institutional arrangements and is subject to strict government regulations. It is compelled to constantly ponder whether it is perceived as a troublemaker by the government. The rules of media regulation and censorship in China are murky and capricious, forcing journalists to constantly play a guessing game of what is allowed by the state. This tendency sometimes even leads to preemptive behavior such as self-censorship. As a result, news media must regularly and tacitly balance a trade-off between newsworthiness and not agitating the government. This structural constraint is where Chinese media deviates the most from its Western counterparts. News media actors can further be classified into three types: government news media, commercial media, and self-media. On one extreme is government news media, which is directly owned by the government or the party, such as the *People's Daily*, *Global Times*, and many other "daily newspapers" owned by various levels of government.⁴⁰

On the other extreme is self-media (*zimeiti* 自媒体), which is the new media actors created by social media, including "verified celebrities, social media influencers, and independent news accounts that produce original content."⁴¹ Commercial media occupies an intermediate position between government news media and self-media. It is most akin to the media studied in the Western context, which mainly follows a market logic. Figure 1 portrays the different types of actors on a spectrum of state embeddedness.

Media selection bias in China

This distinctive political sphere in China leads us to expect structural factors to have an outsized role in shaping how the government accounts report protests on social media. We argue that the government follows the imperatives of stability rather than a market logic. The Chinese government's top priority is to maintain and enhance its legitimacy. Such legitimacy rests upon maintaining a stable society and promoting economic growth. Therefore, the state has a strong tendency to avoid reporting disruptive or violent protests because these events expose the state's vulnerability. By the same logic, the state may be more likely to cover protests if they are against non-state entities and implore the government for help because drawing attention to such instances signals the citizens' trust of the state.⁴² The issue area may be another dimension that separates government and news media accounts. Government accounts may be less likely to report protests organized by social groups that are perceived as threatening to the state, such as those with economic powers or strong mobilizing capacity.

The three types of news media tend to exhibit noted differences according to their embeddedness with the state. Government news media is likely to behave more similarly to government accounts themselves than the two other types of news media with respect to selection bias. On the other end of the spectrum, self-media is not restricted by the "gatekeeper" role of traditional news media and should be more similar to individuals and exhibit the lowest selection bias.

Commercial media, which is most akin to mass media analysed in the Western literature, tends to occupy an intermedia position. It faces an imperative to balance the trade-off between market and state pressure. The commercialization of Chinese media institutions has forced media, even

40 Stockmann and Gallagher 2011.

41 Fang 2022, 4–5.

42 Tang 2016.

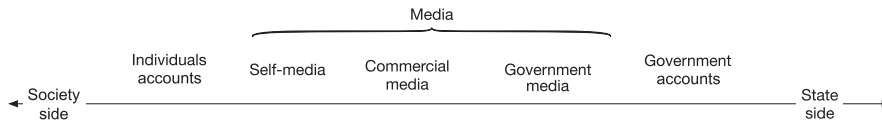


Figure 1. Illustration of Different Types of Actors on Chinese Social Media

Source: the authors

official media, to adopt the logic of the market,⁴³ leading it to favour reporting newsworthy events that can reap enormous public attention and broaden its readership. However, Chinese commercial news media is still subject to strict government regulations.⁴⁴ Therefore, although unconventional tactics – namely, disruptive and violent protests⁴⁵ – are considered to have high news value, only the former may be viewed as acceptable to report in China. The latter – violent protests – are regarded as regime-threatening and are thus unlikely to be covered by the Chinese media. The same holds for the targets of protest action. Although the Western literature suggests that protests addressing issues relevant to the government or legislature often draw greater media attention, Chinese media may eschew reporting on protests that target the government to avoid agitating the state.

Media description bias in China

Considering the circumscribed sphere of the Chinese media, it is probable that the description bias in media portrayal of protest events is amplified in China. On a general level, media actors are likely to construct news content in ways that depoliticize the claims of protesters and marginalize protests that threaten regime legitimacy. Such bias is likely to manifest itself in misrepresentation and differential framing of protests. Specifically, the media may misrepresent key characteristics of protest events by deemphasizing the form of action, the presence of police, and the negative sentiments of the protests. The media may be more likely to portray protests as peaceful rather than violent or disruptive, while less likely to mention government agencies in their descriptions of protests.

Given these considerations, we speculate that, in the China context, the tendency in depoliticizing protest reporting may be especially salient for government accounts on social media. They are likely to describe protests in a more peaceful manner, and with more positive sentiments, less focus on the presence of police, and less mention of government agencies. It can also be predicted that government social media posts exhibit greater description bias than news media. This serves the need of political actors to uphold the image of the state and to foster a positive state–society relationship. In reporting selected events, government accounts can frame these events in such a way that depoliticizes them and deemphasizes the systemic and structural social issues that prompt them. This can result in a portrayal characterized by more *positive* sentiments, *less* violence and *less* policing.

These tendencies partly apply to news media, but to a lesser degree. News media in China is likely to highlight the violent and disruptive behaviors during protests for dramatic effect. But news media also has the disposition to tread a fine line between newsworthiness and conformity to state regulations. Therefore, the media may still neutralize the reporting of disruptive behavior

43 Lei 2016.

44 Stockmann and Gallagher 2011.

45 Violent protests in China take the form of physical conflicts with government officials and/or the police. Disruptive protests often take the form of occupying buildings and land, barricading construction, cutting off power, sitting on roads, and blocking roads or the front entrance of government buildings or organizational offices. Disruptive protests must disrupt public order by occupying public space; protests that are confined to a private space are not considered to be disruptive. See Chen, Xi 2009 and Cai 2010 for discussions.

by framing the protests in a relatively positive tone and decentering the role of the police in resolving the conflict. Whereas government social media posts exhibit greater description bias than news media in general, there is further heterogeneity across different types of news media. Description biases of government news media would be more similar to government accounts, and thus more severe than that of commercial media and self-media. Self-media, on the other hand, is likely to be the most neutral in protest reporting.

One subtle but nonetheless powerful form of description bias involves differential framing. Iyengar⁴⁶ and Smith et al.⁴⁷ identify two styles of protest framing: *episodic* and *thematic*. The former is oriented towards concrete acts that constitute a protest. The latter highlights the general development of the issue and the underlying social tension.

In our context, the style of protest framing is also likely to differ between government accounts and news media. Government accounts are prone to providing a synthesis of government responses to similar protests rather than attending to individual protests in part because of the sheer number of protest events. This practice also showcases the government's systematic efforts and responsiveness in addressing popular grievances without having to explicitly acknowledge the large number of unresolved cases. In this perspective, government accounts are likely to gravitate towards a thematic style.⁴⁸ By contrast, news media may be obligated to cover the details of the protests, although its descriptions are likely biased according to our previous discussions. Hence, we argue that news media is more likely to adopt an episodic style of reporting.

Data and Methods

Data set construction

The protest event data was taken from CASM-China. Zhang and Pan⁴⁹ developed a two-step deep-learning algorithm based on text and image data to identify 118,026 offline collective action events from 9.5 million Weibo posts using 50 general words related to protests. The approach has been extensively validated. Human validations show that CASM-China extracts instances of collective action 10 to 100 times more frequently than newspapers; at the same time, over 90 per cent of protests covered in major Chinese newspapers are captured in CASM-China. CASM-China also covers a wider range of issues than newspapers and does so in a more balanced way. Essentially, underlying this study is the assumption that the CASM data represent protest events in China reasonably well, having overcome the limitations of traditional news media. It constitutes the best available, even if imperfect, source of data on protest events in China.

For each protest in CASM, we created the following variables: geolocation; date; account characteristics (number of followers, followees and posts); issue areas of the protest (land/rural protests, unpaid wages, homeowner property, fraud/scams, environmental, pension/welfare, taxi drivers, medical, education, veterans); protest size; protest target (against state actors including the CCP, against non-state entities such as companies, or against non-state entities but involving the government as a mediator such as imploring the government for help); action form (peaceful, disruptive, or violent); police presence and sentiment of reporting. A detailed description of the variable construction process is in Appendix A (supplementary materials).

⁴⁶ Iyengar 1991.

⁴⁷ Smith et al. 2001.

⁴⁸ On a related note, the recent literature on “consultative authoritarianism” posits that, in contrast to what democratic theorists predict, the Chinese government actively listens to and responds to citizen’s grievances via official platforms built and maintained by the state, such as mayors’ letterboxes and local people’s congresses. See Teets 2013; Su and Meng 2016; Truex 2017. Some of these platforms operate online. Rather than passively answering the complaints (e.g. “message noted”), local officials sometimes address citizens’ complaints by providing or helping them obtain monetary compensation, especially when the threat of collective action looms large. Because of a lack of reliable data, though, it is unclear how often this occurs.

⁴⁹ Zhang and Pan 2019.

Selection biases

To examine the media selection bias for each event in CASM, we identified mentions of the same protests by news media or government accounts on Weibo. We started with the raw data in CASM – the 9.5 million Weibo posts that mentioned at least one of the 50 protest-related words.⁵⁰ We proceeded with two steps: (1) identifying whether an account belonged to the news media or the government; (2) finding mentions of each protest in CASM-China by news media or government accounts.

We first classified each Weibo user into five types as listed in Figure 1. We relied on both the official verification status of Weibo accounts as well as the accounts' usernames to determine whether they belonged to each of the five types. See Appendix B in the supplementary materials for details.

After establishing the account types, we found that out of the total 9.5 million posts that contained protest-related words, 1,111,715 were from news media accounts and 240,591 were from government accounts. To find posts from news media accounts that discussed a particular protest in CASM-China, we applied a first-stage machine classifier to the 1,111,715 media posts,⁵¹ which removed irrelevant posts entirely (e.g. "people are gathering in the plaza for New Year's Eve"; the word "gathering" is also used frequently in protest-related posts). We then kept the media posts that had the same location, at least one overlapping issue and had occurred within a week of the protest to at least one protest in CASM-China,⁵² which resulted in 36,777 media posts. Finally, we manually labelled 2,990 of the 36,777 posts to see if the media posts and the protest that had the same location issue and within a week were actually about the same event (47.8 per cent of them did). We then used the 2,990 posts to train a supervised machine learning algorithm – Random Forest – and apply the trained algorithm to make a final decision on which of the 36,777 media posts were talking about a protest and which were not.⁵³ This ended up with 18,994 media posts that were matched with a protest in CASM-China. These 18,994 media posts were related to 7,694 protests (because some protests were discussed in multiple posts). Furthermore, the composition of the 18,994 media posts is as the following: 237 government news media accounts posted 2,031 posts; 612 *zimeiti* accounts posted 2,591 posts; and 2,627 commercial accounts posted 14,372 posts. Hence, commercial media is the most popular media type, and it represented the majority of the media posts.

To find posts by government accounts that discussed a specific protest in CASM-China, we followed similar steps and obtained the 2,896 government posts. Because this time the number is small, we therefore relied on research assistants to read all 2,896 posts in detail and find 810 posts that were indeed about a particular protest or a group of similar protests in a specific city. These 810 posts were related to 530 protests. We ended up using these 530 protest events that are verified by humans as the matched protests in government accounts. Figure 2 is an illustrative chart summarizing our process of constructing mentions of protests by news media and the government.

It remained possible that if we could not find media or government posts about a protest in CASM-China, it was because these discussions did not use the 50 protest-related words at all such that they did not enter our raw data. To mitigate this possibility, we randomly selected 30 unmatched CASM posts and checked whether we could find mentions of these protests beyond the 9.5 million posts. Only three out of the 30, or 10 per cent, were *not* in the 9.5 million Weibo

50 About 3.85% of protests in CASM-China included multiple locations. We discarded these events from the beginning.

51 We used the publicly available CASM's first stage classifier.

52 We carried out sensitivity analyses relaxing the one-week timeframe to two, four and eight weeks, which led to similar results.

53 The Random Forest model achieved a precision score of 0.89 and a recall score of 0.76 through fivefold cross-validation of the human-labelled 2,990 matched posts.

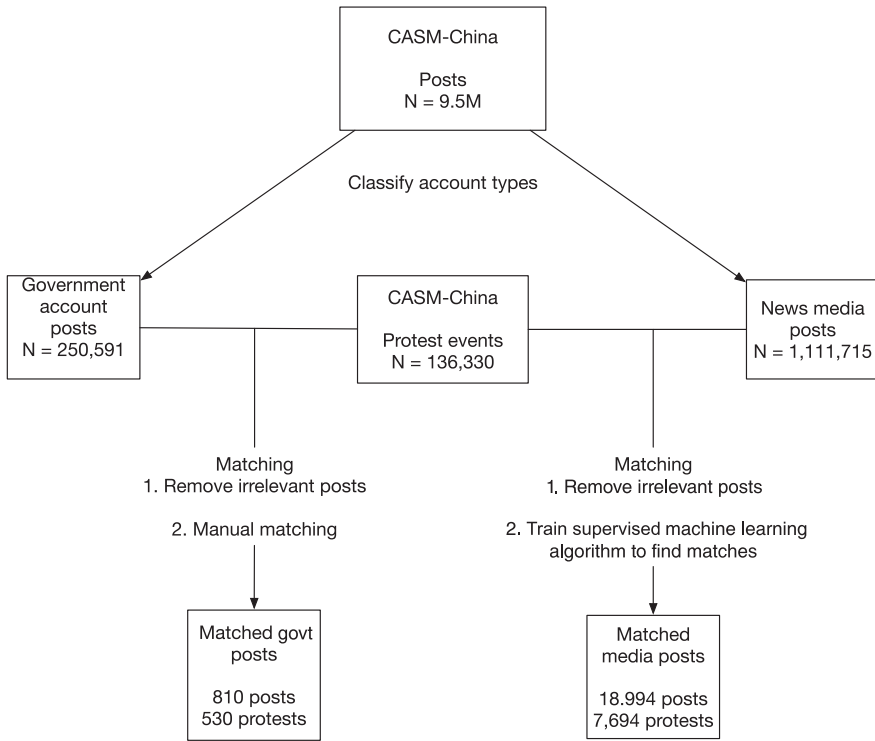


Figure 2. Post Matching Flowchart
Source: the authors

posts, suggesting that the omission bias due to search queries is small. One additional concern was that this 10 per cent of posts utilized different vocabulary to describe protests compared to the remaining 90 per cent, potentially confounding our results on description biases. However, we observed no such discrepancies. Consequently, we believe our approach was robust against omission biases.⁵⁴

Next, we studied factors that predict whether and how often a protest in CASM was mentioned by news media or government accounts. Specifically, we ran logistic regression models to test which factors explained whether a protest event in CASM was covered by news media or government accounts. We first estimated models by two broad categories of accounts (government vs. news media). We then further distinguished between three types of news media. We also fit quasi-Poisson regression to model the frequency of reporting. Quasi-Poisson regression models separately estimate the variance parameter and thus allow the variance to be greater than the mean, which relaxes the assumption of regular Poisson regression that assumes the mean and the variance of the dependent variable is the same. Both the logistic regression and quasi-Poisson regression used event characteristics and user characteristics as the explanatory variables (based on CASM posts).⁵⁵ In all the models used in this study, we included the provinces and year-fixed effects to adjust for stable unmeasured provincial characteristics and for macro-sociopolitical changes over time.

54 All of the 10 per cent were because they used homophones or inserted irrelevant characters to evade censorship, which were not included in CASM-China’s keyword lists. For instance, when people write 游行, they can write it as 游 | 行 to fool the censorship algorithm.

55 User characteristics were taken directly from meta information from Weibo, including the user’s number of followers and followees as well as the number of total posts at the time of data collection (June 2020).

Analysis of description bias

To study the media description bias, we compared the descriptions of protests by individuals with those of the news media or the government accounts along a series of characteristics. These analyses were based on only the CASM events that were matched with news media or government accounts. The same method used to generate protest characteristics in the overall CASM data set was used to construct characteristic variables for protests covered by news media and government accounts: namely, action forms, sentiment, presence of police, and mention of the state. We first estimated models by two broad categories of accounts (government vs. news media). We then further distinguished between the three types of news media.

Results

Descriptive statistics

We obtained data on over 122,631 protests from CASM. Among them, 530 protests (or 0.4 per cent) were covered by Weibo accounts affiliated with local governments and 7,694 protests (or 6.3 per cent) were reported by news media accounts on social media. These results show significantly fewer reports by the government compared to the news media. This gap may be partly explained by the greater number of news media accounts than government accounts on Weibo. In addition, the government accounts tended to report fewer protests for legitimacy purposes and for the sake of reserving space for propaganda-related materials.

Table 1 presents the summary statistics of the variables included in the analysis. As expected, the government and news media accounts had a greater number of followers and followings than individual accounts did. They were also more likely to verify themselves on Weibo. Regarding the issues, the government posts were more likely to report on protests related to unpaid wages, while the news media posts' tendency to report on protests is similar to that of individuals. As for the coverage of the protests' form of action, the government posts were significantly more likely to report on peaceful protests, while news media posts gravitated towards more violent protests. Regarding targets, government posts were more likely to cover protests directed at companies, especially when the government served as the mediator, and news media posts were more likely to cover protests against either companies or the government. These results should be interpreted with caution because they were not adjusted for other factors. We provide a more systematic analysis in a regression framework below.

Media selection biases

Table 2 shows the regression results using two-way fixed-effect regression (at province and year level), and the standard errors are clustered at the provincial level.⁵⁶ The first two columns display the results from logistic and quasi-Poisson regressions that predicted whether and how many times a protest was covered by news media accounts based on user-level and protest-level covariates. The third and fourth columns show the results of logistic and quasi-Poisson regression that predicted whether and how many times a protest was covered by government accounts. The analysis was based on complete cases after dropping missing values. The results were largely similar regardless of whether we used dummy or count outcome variable measures (i.e. using logistic or quasi-Poisson regression).

The results show that the characteristics of the user and the protest matter in the selection process. Posts on protests by more influential and/or popular users were more likely to be reported by

⁵⁶ We did not further cluster standard errors at the year level because there are not many years, and clustering with a short panel will result in finite-sample biases. See Angrist and Pischke 2008 for discussions.

Table 1. Summary Statistics

	Individual N=122631	News media N=7694	Government N=530
Size (log)	3.03 (1.12)	3.21 (1.19)	3.33 (1.24)
Number of posts (log)	6.32 (2.51)	7.81 (2.18)	7.17 (2.47)
Number of followers (log)	5.85 (2.99)	8.53 (3.68)	7.52 (3.93)
Number of followees (log)	5.49 (1.44)	6.05 (1.22)	5.81 (1.28)
Verified	14.2%	34.7%	19.1%
Issue:			
Land grabs	0.23 (0.42)	0.23 (0.42)	0.16 (0.36)
Unpaid wages	0.30 (0.46)	0.37 (0.48)	0.59 (0.49)
Homeowners	0.26 (0.44)	0.30 (0.46)	0.15 (0.35)
Fraud	0.08 (0.26)	0.08 (0.27)	0.06 (0.24)
Environment	0.05 (0.21)	0.06 (0.23)	0.08 (0.27)
Pension	0.04 (0.19)	0.05 (0.21)	0.04 (0.20)
Taxi	0.04 (0.20)	0.06 (0.24)	0.04 (0.20)
Medical	0.12 (0.32)	0.15 (0.36)	0.14 (0.35)
Education	0.08 (0.27)	0.11 (0.32)	0.10 (0.31)
Veteran	0.01 (0.07)	0.01 (0.07)	0.00 (0.00)
Action form:			
Peaceful	37.0%	53.6%	72.8%
Disruptive	16.4%	11.6%	5.45%
Violent	46.6%	34.8%	21.8%
Police presence:			
No police presence	77.8%	75.9%	71.3%
Police presence	22.2%	24.1%	28.7%
Target:			
Company	40.2%	45.3%	43.2%
Company; government as mediator	20.5%	22.0%	27.9%
Government	39.3%	32.7%	28.9%
Mention state	50%	47%	55%
Sentiment	-0.03 (0.11)	-0.02 (0.10)	-0.02 (0.08)

Source: the authors

news media and governments. However, if the protest was posted by a user who followed a lot of social media users, it was less likely to be picked up by the news media, perhaps because following many people (instead of being followed by many people) signalled the lower status of the user.

The size of the protest increased the probability of selection by both news media and government accounts. The presence of police increased the probability of selection only for news media; for government accounts, its impact is positive but not statistically significant. This finding is consistent to what has been well established in the literature regarding protests in Western countries: news media tends to report protests with larger sizes and with police involvement. The finding that the Chinese

Table 2. Probability of Reporting a Protest Event by News Media or Government Accounts, Based on Two-Way Fixed-Effect Regression at Province and Year Level with Clustered Standard Errors at Province Level

Model	News media	News media	Government	Government
	Logit	Quasi-Poisson	Logit	Quasi-Poisson
Number of posts (log)	0.0496** (0.0178)	0.0561*** (0.0153)	-0.0031 (0.0386)	-0.0013 (0.0378)
Number of followers (log)	0.2557*** (0.0125)	0.2007*** (0.0111)	0.2048*** (0.0293)	0.1970*** (0.0279)
Number of followees (log)	-0.0811** (0.0287)	-0.0561* (0.0238)	-0.0531 (0.0542)	-0.0505 (0.0531)
Verified	-0.0737 (0.0647)	-0.0152 (0.0527)	-0.5420*** (0.1457)	-0.5182** (0.1417)
Size (log)	0.0989*** (0.0157)	0.0821*** (0.0129)	0.1562+ (0.0833)	0.1518+ (0.0803)
Issue:				
Land grabs	0.1278 (0.0665)	0.1034 (0.0547)	-0.1668 (0.3261)	-0.1662 (0.3210)
Unpaid wages	0.3527*** (0.0575)	0.2851*** (0.0478)	0.8902*** (0.1859)	0.8684*** (0.1809)
Homeowners	-0.0300 (0.0573)	-0.0182 (0.0480)	-0.8337*** (0.2474)	-0.8223** (0.2453)
Fraud	-0.1337* (0.0549)	-0.1062* (0.0465)	-0.3599 (0.2547)	-0.3499 (0.2478)
Environment	0.0553 (0.0907)	0.0488 (0.0768)	0.9107** (0.2784)	0.8917** (0.2666)
Pension	0.1613* (0.0779)	0.1205 (0.0622)	-0.1661 (0.3111)	-0.1621 (0.3026)
Taxi	0.1061 (0.0611)	0.0730 (0.0479)	-0.0792 (0.2442)	-0.0791 (0.2367)
Medical	0.0172 (0.0554)	0.0171 (0.0445)	0.1434 (0.1750)	0.1416 (0.1690)
Education	0.0808 (0.0763)	0.0580 (0.0616)	0.2829 (0.2309)	0.2749 (0.2222)
Veteran	0.1018 (0.1809)	0.0897 (0.1459)	-12.38*** (0.1370)	-13.10*** (0.1308)
Police presence	0.1360* (0.0578)	0.1136* (0.0481)	0.2991 (0.1977)	0.2919 (0.1932)

(Continued)

Table 2. (Continued.)

Model	News media	News media	Government	Government
	Logit	Quasi-Poisson	Logit	Quasi-Poisson
Action form:				
Disruptive	0.0168 (0.0748)	0.0241 (0.0611)	-0.6680** (0.2149)	-0.6523** (0.2110)
Violent	-0.0666 (0.0582)	-0.0535 (0.0483)	-0.2234* (0.1218)	-0.2181* (0.1179)
Target:				
Company; government as mediator	-0.0172 (0.0520)	-0.0244 (0.0429)	0.0691 (0.2345)	0.0626 (0.2272)
Government	-0.1990*** (0.0600)	-0.1769** (0.0507)	-0.1703 (0.1706)	-0.1690 (0.1663)
Sentiment	0.1616 (0.1354)	0.1466 (0.1148)	0.6746 (0.5697)	0.6628 (0.5584)
Observations	56,376	56,376	55,992	55,992

Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, +: 0.1

Source: the authors

government is also more likely to report it might be surprising for some. Protests that involved many participants were considered newsworthy and were more widely known. Deliberately neglecting such events could challenge the credibility and authority of news media and government Weibo accounts, and prompt citizens to actively seek out information that the government intends to hide.⁵⁷ Instead, the government tends to adopt a strategy of reporting on large-scale protests or protests with police presence but framing them in a more positive, depoliticized way, as detailed in the next section.

Regarding issues, both the news media and government accounts were more likely to report on protests caused by unpaid wages. Other than this similarity, the focus of the reports delivered by the news media and government diverged. The news media was more likely to report on protests associated with unpaid wages and pensions, but refrain from reporting on fraud. Government accounts were also more likely to report on protests organized by workers on unpaid wages. They were also prone to reporting on protests organized by residents involving environmental issues (e.g. building factories near lands). In the meantime, government accounts avoided reporting on property rights demonstrations organized by the urban middle class (i.e. homeowners). The government was also less likely to report on protests by veterans, which may be explained by the high organizational capacity and tight relationships of veterans that transcend local boundaries; these types of protests frequently involve thousands of people from around the country. These two groups carry strong organizational capacity and resources: homeowners are frequently organized by the urban middle classes who have the necessary monetary resources and knowledge for organizing collective action. Veterans have strong ties and, because they spread across the country, they have the ability to mobilize across geographic boundaries, posing a significant threat to the regime.⁵⁸

⁵⁷ Francisco 2005; Roberts 2018.

⁵⁸ Note that our discussions on how media biases vary by issues are not exhaustive, in part because this study focuses on how media biases vary by media type. Little research has been conducted on media bias by issue, particularly in China. Thus, our findings pave the way for more extensive future investigation.

Regarding action forms, there was no statistically significant difference between news media and individual accounts. This observation contrasts sharply with findings from similar research on Western societies, which demonstrate that the Western news media reports disproportionately disruptive or violent protests. As expected, the government accounts shied away from reporting on escalated protests, such as violent and disruptive ones.

Regarding targets, the news media accounts avoided reporting on sensitive topics that targeted the government. This should come as no surprise since they need to balance the imperatives of newsworthiness and political sensitivity. In comparison, the government accounts did not seem to shun protests against the government itself. This observation may seem surprising at first. However, as discussed below, this finding should be understood in the context of the description bias. In other words, government accounts did not shy away from protests against the state but framed such events in ways that depoliticized and desensitized the issues surrounding the demonstrations.

Overall, the results highlight some broad similarities and important differences between news media and government reports of protests. In terms of the absolute number of protests reported, the government covered significantly fewer events than the news media. This could have resulted from government accounts having to strike a balance between different topics, the lower number of government accounts than news media accounts on Weibo, and/or the fewer posts published by the government accounts than the news media accounts. As for the factors shaping the coverage of protest events, both the news media and the government selected the issue area, action form, target, size and police presence, albeit in different ways.

The media and the government's priorities regarding issue areas were markedly different. The news media selected for newsworthiness, but at the same time self-censored when presented with events that targeted the government. Unlike their Western counterparts, news media accounts in China did not tend to cover disruptive or violent protests to a greater extent. This could also be explained as a result of self-censorship. The government accounts marginalized disruptive and violent protests, and protests organized by the urban middle classes or by veterans.

Table 3 further splits news media into three types: government news media, commercial media, and self-media. The results point to significant differences across the three types of media. Notably, government news media (such as the *People's Daily*) is much more similar to government accounts (in Table 2) than to self-media. For instance, government news media, similar to government accounts, was less likely to report disruptive and violent protests, whereas self-media was more likely to report disruptive protests, which follows naturally from a market logic. On the other hand, self-media did not report protests targeting the state to balance the risk and shield itself, whereas government news media deviates from this pattern. The scales of coefficient estimates of the commercial media are between the government news media and self-media, suggesting that commercial media is more restricted than self-media but also distinguishes itself from government news media (i.e. it does not entirely serve as the state's mouthpiece).

We carried out an additional analysis: we further filtered posts by passerby individuals from all the individual posts.⁵⁹ We found that most individuals are indeed protesters: around 2.5 per cent of posts included one of these words. We removed these 2.5 per cent of posts and reran our analysis. The results, which are consistent with the main text, are presented in Appendix Table E.1 (supplementary materials).

We also found that around 30 per cent of protests had two issues. This may be explained by the lack of precision in CASM's algorithm to classify the issue, which is based on the dictionary method and did not use the most advanced machine learning techniques. It may also be because some protests genuinely span multiple issues (for example, taxi driver's protests may also relate to unpaid

59 The criteria were whether the posts had any of the following words: *weiguan* 围观 (witness), *pangguan* 旁观 (witness), *lugu* 路过 (to pass by), *shangban lushang* 上班路上 (on the way to work).

Table 3. Probability of Reporting a Protest Event by Government, Commercial and Self-media Accounts, Based on Two-Way Fixed-Effect Ordinary Least Squares (OLS) at Province and Year Level with Clustered Standard Errors at Province Level

Model	Government news media	Commercial media	Self-media (<i>zimeiti</i>)
Number of posts (log)	0.0806** (0.0288)	0.0470* (0.0196)	0.0193 (0.0323)
Number of followers (log)	0.2688*** (0.0222)	0.2245*** (0.0135)	0.2892*** (0.0177)
Number of followees (log)	-0.0816* (0.0385)	-0.0154 (0.0273)	-0.1788*** (0.0528)
Verified	-0.6358*** (0.1246)	-0.3445*** (0.0752)	0.8073*** (0.1273)
Size (log)	0.1085** (0.0406)	0.0844*** (0.0185)	0.1347*** (0.0230)
Issue:			
Land grabs	0.1969 (0.1408)	0.0825 (0.0690)	0.2147 (0.1152)
Unpaid wages	0.7263*** (0.1150)	0.2290** (0.0736)	0.2536** (0.0878)
Homeowners	-0.3749** (0.1396)	-0.0638 (0.0662)	0.2231** (0.0711)
Fraud	-0.3578* (0.1771)	-0.0830 (0.0832)	-0.0474 (0.1255)
Environment	0.1534 (0.2248)	0.0025 (0.1055)	0.0492 (0.1711)
Pension	-0.0177 (0.1807)	0.2545** (0.0869)	0.1132 (0.1053)
Taxi	0.3531*** (0.1057)	0.0410 (0.0682)	0.0548 (0.1616)
Medical	0.0029 (0.1050)	0.0364 (0.0711)	-0.0216 (0.1259)
Education	0.0417 (0.2006)	0.1251 (0.0927)	0.0026 (0.1079)
Veteran	0.6421 (0.3506)	-0.2026 (0.2501)	0.0896 (0.4594)
Police presence	0.3224*** (0.0853)	0.0662 (0.0712)	0.0977 (0.1015)

(Continued)

Table 3. (Continued.)

Model	Government news media	Commercial media	Self-media (zimeiti)
Action form:			
Disruptive	-0.4519*** (0.1342)	0.0670 (0.0897)	0.2234** (0.0817)
Violent	-0.3041** (0.0940)	-0.0812 (0.0672)	0.1617 (0.1020)
Target:			
Company; government as mediator	0.2241* (0.0988)	-0.0199 (0.0695)	-0.1302 (0.1005)
Government	0.0056 (0.0964)	-0.0935 (0.0705)	-0.5679*** (0.1347)
Sentiment	0.1585 (0.3076)	0.0248 (0.1824)	0.5497 (0.2822)
Observations	52,528	54,099	52,219

Signif. Codes: ***: 0.001, **: 0.01, *: 0.05
Source: the authors

wages). It is not straightforward to separate the proportion of the two types, though. We ran regressions to include the number of issues of each protest as a control variable in Appendix Table E.2 (supplementary materials). We found that the protests with only one issue or multiple issues did not have a statistical difference in their probability of being reported by news media and government accounts.⁶⁰

Media reporting biases

When protests were covered, how did the government and news media's reports of them differ from the descriptions from individuals? For each protest that was covered by government or news media accounts, the description of event characteristics (action form, sentiment, police presence and mention of the state) was modelled based on a dichotomous variable indicating whether the description was made by the government accounts (or news media), coded as 1, versus by individuals (coded as 0), while controlling for the other covariates. For action forms in particular, a multinomial logit model was used to distinguish between disruptive and violent protests. For other dependent variables, linear regressions were used. Protest-level fixed effects were used. This way any variable stable at the protest level was effectively adjusted for (e.g. the location, the date and the issue the protest was about).

Panel A of Table 4 shows the results of comparing the news media and individuals' descriptions of the same event. News media accounts were no more likely to portray the protests as disruptive. However, they were significantly more likely to portray the protests as violent when compared with the individuals' descriptions. The news media reports were also more likely to exhibit more positive sentiments.⁶¹ They were less likely to mention the police presence, but they were more likely to mention the government.

⁶⁰ We thank an anonymous reviewer for making these suggestions.

⁶¹ To check the robustness of the sentiment measures, we applied an open-source Chinese sentiment classification algorithm, PaddlePaddle, developed by Baidu using deep-learning algorithms. The estimated coefficient was 0.0014 with a

Table 4. Description Bias in the News Media and Government Descriptions

Outcome Level	Action form Disruptive	Action form Violent	Sentiment	Police presence	Mention state
Panel A: Description biases between the news media and protesters' descriptions of the same event					
News media	0.0091 (0.0052)	0.0838*** (0.3304)	0.0070*** (0.0007)	-0.0844*** (0.0031)	0.0433*** (0.0038)
Observations	37,552	37,552	36,848	37,552	37,552
R ²	0.84689	0.82301	0.60684	0.76800	0.73134
Panel B: Description biases between the government and protesters' descriptions of the same event					
Government	-0.9182*** (0.0319)	-1.2016** (0.0226)	0.0128*** (0.0027)	-0.1475*** (0.0169)	0.0775*** (0.0229)
Observations	1,600	1,600	1,577	1,577	1,600
R ²	0.82845	0.82819	0.51419	0.55466	0.57564
Panel C: Statistical test between two coefficients (government - news media)					
Bootstrapping (over 10,000 samples) was used to calculate the p-values.					
Difference in means	-1.0268	-1.4091	0.0058	-0.0649	0.0346
p-value	0.0584 ⁺	0.0101*	0.0581 ⁺	0.0023**	0.1228

One-way (protest) standard-errors in parentheses

Signif. Codes: ***: 0.001, **: 0.01, *: 0.05, +: 0.1

Source: the authors

Panel B of [Table 4](#) shows the results of a comparison between the government's and individuals' descriptions of the same event. The government reports were less likely to portray the protests as either disruptive or violent when compared to individuals' descriptions. Government descriptions were more positive in sentiment and less likely to mention the police presence than individuals' descriptions. Again, this observation may be explained by the fact that the government often discusses its role in the resolution of protest grievances, which is discussed in detail in the next section.

It is important to note that for sentiment, police presence and mentions of the state, the difference between news media reports and individuals' descriptions was smaller than the difference between the government's reports and individuals' reports (as measured by the magnitude of the coefficients). We used bootstrap resampling to conduct statistical tests for the difference between two regression coefficients across two samples.⁶² As shown in Panel C of [Table 4](#), we found that there were statistically significant differences between the two sets of coefficients (except mentions of the government). This confirmed the prediction that the government's description bias is greater than that of the media. Specifically, the government portrayed protests as less violent, with less police involvement and more positive than the individuals' own accounts. The news media's portrayals of the protests were also biased on certain dimensions, but less consistently and to a lesser extent than government accounts.

[Table 5](#) further shows the differences in reporting biases by the three types of news media: government news media, commercial media and self-media. Again, the reporting biases of government news media are more similar to that of government accounts, compared to the other two types of media. For instance, government news media, similar to government accounts, framed the protests as less disruptive, compared with individuals' descriptions. On the other end, self-media is more similar to individuals. The only area where self-media and individuals diverges is in their description of police presence: self-media was less likely than individuals to mention police at the scene. For other dimensions of description, there were no statistically significant differences between self-media and individuals. Last, the majority of news media accounts were still commercial media (see the number of observations in [Table 5](#)), and commercial media's description biases conveyed exactly the same story as we have seen for news media in general (Panel A, [Table 4](#)). In general, [Table 5](#) again finds support for the distinction across the three types of news media in protest portrayal.

Episodic versus thematic reporting

Finally, we compared the news media and government accounts in their style of reporting. We hypothesized that the state tends to cover protests with a thematic style, while news media gives more attention to details of protests in an episodic style (although both are subject to description biases).

We first offer some examples of episodic versus thematic reporting of the same CASM event. Here we discuss an exemplary event – a protest against the construction of a waste incineration plant in the Zhongtai subdistrict of Yuhang District, Hangzhou City.⁶³ In this incident, the protesters demonstrated against the construction of a waste incineration plant near their homes. The protest took place on 10 May 2014, with more than 5,000 participants. There was a lot of social media

standard error of 0.0009. The p-value was smaller than 0.001. Hence, the finding that the news media tended to describe the same event using more positive tones than the individuals is robust.

62 We thank a reviewer for suggesting the use of bootstrap for comparing models.

63 We chose this event because it is one of the most discussed events on CASM-China and it has also attracted a lot of domestic and international media attention; “Zhejiang Hangzhou kangyi laji fadianzhan yin chongtu jin sishi shang” (Protest against incineration construction in Hang Zhou, Zhejiang has caused 40 injuries), *BBC News*, 11 March 2014, www.bbc.com/zhongwen/trad/china/2014/05/140511_china_hangzhou_environment_protest, accessed 9 June 2023. The other two posts were chosen because they provide details to illustrate different styles of reporting.

Table 5. Difference between the Three Types of Media and Individuals' Reports; Individual as Reference Group

Outcome Level	Action form Disruptive	Action form Violent	Sentiment	Police presence	Mention state
Government news media	-0.2796* (0.1169)	0.0441 (0.0658)	0.0029 (0.0021)	-0.0901*** (0.0099)	0.0950*** (0.0131)
Observations	4,062	3,983	3,983	4,062	4,062
R ²	0.7724	0.4603	0.54752	0.76140	0.64836
Commercial media	0.0097 (0.0375)	0.0952*** (0.0261)	0.0128*** (0.0010)	-0.0906*** (0.0037)	0.0415*** (0.0044)
Observations	28,308	28,308	27,802	28,308	28,308
R ²	0.7886	0.6347	0.58438	0.76347	0.72234
Self-media	0.1069 0.0739	0.0552 0.0021	0.0029 (0.0017)	-0.0463*** (0.0071)	0.0131 (0.0079)
Observations	5,182	5,182	5,063	5,182	5,182
R ²	0.8342	0.72	0.74941	0.79942	0.83852

Clustered (event) standard-errors in parentheses

Signif. Codes: ***: 0.001, **: 0.01, *: 0.05

Source: the authors

discussion on this protest. One participant wrote in his Weibo post (English translation comes first and the original post in Chinese follows):

The collective action event in Hangzhou has escalated. Every street in the Yuhang District is full of protesters. We are only protesting for our future. We firmly oppose the construction of a waste incineration plant in Zhongtai subdistrict! We are engaging in such action for our children, for our living environment and for our homeland! (杭州群体性事件已经升级。余杭大街满是抗议活动，为的只是我们的将来。坚决抵制中泰建造垃圾焚电厂，为了我们的子孙后代，为了我们生活的环境，为了我们的家园。)

Our close reading suggests that the news media reported on the same event primarily in an episodic style. Below is a post by the *New Beijing News* (*Xinjing bao* 新京报), an influential local traditional media outlet that has successfully established itself as a popular nationwide social media account (with over 46 million followers on Weibo, as of August 2022). The owner of the *New Beijing News* is the propaganda department of the Beijing municipal committee of the CCP. Therefore, the *New Beijing News* is classified as a government news media under our classification scheme. The report by the *New Beijing News* provided four of the classic “five Ws” used in news reporting (“who, what, when, and where”) but did not mention why the people were protesting or any of the protesters’ perspectives. The *New Beijing News* also made no further comment on its attitude towards the protest.

In response to the protests against the construction of a waste incineration plant in Zhongtai subdistrict, Yuhang District, the Hangzhou Public Security Bureau announced today that under the incitement of a small group of criminals, a group of people has gathered around

Zhongtai subdistrict. Some criminals smashed cars and assaulted police officers as well as pedestrian bystanders. 53 people have been arrested. Another seven have been placed under administrative detention because they spread false information about the protest online. (关于余杭中泰垃圾焚烧厂抗议事件，杭州公安局今天公告在少数不法分子煽动蛊惑下余杭中泰及附近地区人员规模性聚集一些不法分子打砸损坏车辆围攻殴打执法民警和无辜群众已刑拘53人另有7人在网上捏造谣言散布虚假信息被行政拘留)

The government reported on the same event using a thematic style. Below is a quote from a post by Hotline 12355 (12355 *qingshaonian rexian* 青少年热线), the official Weibo account of the Communist Youth League (*gongqingtuan* 共青团) in Yuhang District, Hangzhou City. The post from this government account did not mention much detail of the protest at all. Specifically, it did not provide details on who was protesting, the exact name of the construction site (only the locality), or the date of the protest. The government post mainly used the protest as an example to discuss the root causes of this type of protest and how the government should prevent these types of protests and rebuild trust by using effective and transparent communication. Furthermore, it is evident that the blame for the protest was placed on the local government.

The construction of a waste incineration plant in Yuhang, Zhejiang, triggered popular grievances and evolved into a violent incident involving the smashing of police cars and the assault of police officers. The key reason behind the event is the loss of trust in local government in environmental protection. The local government must take actions to rebuild people's trust in its capacity to protect the environment. It also needs to facilitate effective communication and transparency. This will avoid the dilemmas of construction projects being interrupted by protests. (浙江余杭建垃圾焚烧厂引发民众不满并演变为烧警车袭击警员的暴力事件。地方政府环保信用缺失是造成此类事件的根源。地方政府只有通过实际行动重建其环保信用，才能谈得上有效的沟通和透明。避免一建就闹一闹就停的窘境。)

These examples illustrate different reporting styles by the government and news media on the same protest event. The government mainly resorted to a thematic style and omitted the details, whereas the news media provided greater details in an episodic style. To provide a more systematic analysis of this diverging pattern, we identified the top 25 words used by individuals (measured by word frequency). Typically, individuals' descriptions of the protests included more details of the protests. We then calculated the corresponding ranking of these top 25 words in the news media and government posts. If the government tends to use a thematic style, the words frequently used by them would differ from those used by individuals.

Table 6 shows that the frequency of these words (i.e. the rankings) in posts by individuals and by news media accounts was fairly similar. However, a notable gap emerged between individual and government accounts. The words most frequently used by individuals, such as "police," "real estate developers," "protest banners," and "besiege," were not frequently taken up by government accounts. Instead, the government frequently used words such as "court" (1st), "situation" (2nd), "begin/expand" (3rd), "company" (4th), "legal case" (5th), and "protect" (6th). These words are, for the most part, related to the government's efforts to address grievances and to protect the rights of protesters through legal procedures. These results provide additional evidence that the government's reports are more "thematic," while the news media reports are more "episodic" and more similar to individuals' descriptions. To visualize these patterns, we also plot the content of Table 6 in Figure E.1 in the Appendix (supplementary materials) for interested readers.

Table 6. Top Words Ranked by Frequency by Weibo Posts from Individuals, News Media and Government Accounts

Rank	Individual	News media	Government
Word:			
<i>jingcha</i> 警察 (police)	1	3	21
<i>kaifashang</i> 开发商 (real-estate developers)	2	1	22
<i>hengfu</i> 横幅 (protest banners)	3	2	19
<i>weidu</i> 围堵 (besiege)	4	5	20
<i>xiaoqu</i> 小区 (residential complex)	5	4	24
<i>shenbian</i> 身边 (nearby)	6	7	25
<i>yuangong</i> 员工 (employees)	7	6	23
<i>qingkuang</i> 情况 (situation)	8	8	2
<i>fayuan</i> 法院 (court)	9	12	1
<i>qiye</i> 企业 (company)	10	9	5
<i>gongzuo</i> 工作 (work)	11	10	9
<i>weihu</i> 维护 (protect)	12	18	7
<i>yifa</i> 依法 (according to law)	13	13	16
<i>zhifu</i> 支付 (payment)	14	11	4
<i>fafang</i> 发放 (distribute)	15	16	12
<i>laodong</i> 劳动 (labour)	16	14	15
<i>huodong</i> 活动 (activity)	17	19	14
<i>anjian</i> 案件 (legal case)	18	17	6
<i>baozhang</i> 保障 (social security)	19	15	11
<i>hefaquanyi</i> 合法权益 (legal rights)	20	20	8
<i>kaizhan</i> 开展 (begin/expand)	21	21	3
<i>jizhong</i> 集中 (focus)	22	25	17
<i>jiancha</i> 检查 (investigation)	23	22	13
<i>zhuanxiang</i> 专项 (special project)	24	24	10
<i>falüyuanzhu</i> 法律援助 (legal assistance)	25	23	18

Source: the authors

Conclusions

The present study advances the understanding of media biases in protest reporting in China. In doing so, it sheds light on the complex interplay between the state, the media and society in contentious politics. Past research on how the media report protests has focused on the traditional mass media (i.e. newspapers or television) and is overwhelmingly centered on Western societies. There is also scarce research on media reporting biases on social media platforms and on the differences between government reports and news media reports. The rise of social media in recent decades has provided ample opportunities for the traditional mass media and the government to have a strong presence in virtual spaces. This article is among the first to examine the reporting of protest events by news media and government social media accounts in China, producing results with broad relevance.

We examined two aspects of media reporting biases, namely selection bias and description bias, and found evidence for both. With respect to selection bias, both news media and government accounts were selective in their coverage of protests but in distinct ways. Unlike their Western

counterparts, the Chinese news media engaged in a delicate balancing act between a market logic and structural regulation. They tended to move away from reporting protests targeting the government, but gravitated towards protests organized by disadvantaged groups. The government accounts, on the other hand, shied away from reporting violent protests, as well as protests by the urban middle class and veterans – the two groups with a particularly strong mobilizing capacity that can potentially transcend local areas.

Additionally, both the government and news media reports were subject to descriptive biases, and such biases were more pronounced on government accounts than news media accounts. The Chinese news media sought to strike a delicate balance between a market logic and state restrictions, which manifested in a high degree of selection bias and a moderate level of description bias. The Chinese government accounts, in contrast, engaged in a comparatively moderate degree of selection bias and a high degree of description bias. They tended to report on visible protest events of broad relevance with less consideration of their political sensitivity. It did, however, construe the events in ways that depoliticized individuals' claims and enhanced the state's legitimacy. This was achieved by portraying protests with more *positive* sentiments, less *violence* and less *policing*. Moreover, government coverage of protest events also exhibited a more thematic orientation than the news media coverage.

Embeddedness within the state further stratifies news media sources, resulting in differential selection and description bias among various types of news media. Government news media behaves more similarly to government accounts than two other types of news media. Self-media, on the other hand, is neither bound by the "gatekeeper" position of traditional news media nor directly supervised by the Chinese government. Indeed, we found that self-media had the least amount of selection and description bias: it was most comparable to individuals on Weibo. Commercial media, which more closely resembles the mass media discussed in the Western context but is nevertheless regulated by an authoritarian state, tends to occupy an intermediary position in the level of selection and description bias, between self-media and government news media.

This study has several limitations, which open the door to future research. Notably, the data set was generated by machine prediction. Although the predictions had a high level of accuracy, the errors may have still carried over to the regression model estimates. In the methodological literature, there has been some recent progress in the discussion of how machine prediction errors should be accounted for in next-step regression models,⁶⁴ but these studies have not considered cases in which there are many variables in the regression being predicted, as was the case in this article. Furthermore, the collection of government and news media accounts was based on Weibo posts using 50 protest-related words. An ideal design would sample all government or media accounts, but currently there is no combined list of these accounts. Additionally, censorship may also bias the results, although government posts may theoretically carry a low risk of being deleted by the propaganda machine. Despite such limitations, we believe the benefits of our data outweigh its drawbacks.

Finally, we did not have data for traditional newspapers in their original printed format. It requires a separate data set to evaluate the extent to which our results generalize to printed newspaper articles. A comparison between traditional newspaper and social media is beyond the scope of this study. We anticipate, however, that the news media's reporting in traditional format is subject to additional control. Each newspaper report goes through comprehensive reviews (*sanshen sanjiao* 三审三校) before appearing in print. However, there has been no report showing that such a rigorous ex-ante review process has been used by traditional media on social media. Future study is required to empirically validate this hypothesis.

The present research has broader implications for the study of social movements both within and outside of China. Recent literature has discussed why the Chinese state has allowed some space for protest reporting in China. This study contributes to this burgeoning literature by providing a more nuanced understanding of how the state strategically selects and frames protests in a way that serves

64 Fong and Tyler 2020.

its own agenda. This study also contributes to the study of media biases in reporting on social movements more broadly. Most studies on media bias in social movements coverage are conducted in the Western world, where a relatively more open media environment is coupled with strong market incentives. Reports of protests in authoritarian regimes have received scant attention. This article enriches the scholarly understanding of how the news media operates in an authoritarian regime by revealing how the news media in China strategically reports on protests to balance commercial interests while also accommodating state restrictions. These results of differential patterns across media characterized by varying state embeddedness will be helpful for future research investigating the complex interactions between social movements, media and the state in a variety of regimes.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0305741023001042>.

Acknowledgements. The authors would like to thank Linda Cheng and Yongxin Ke for their excellent research assistance. We acknowledge support from the Weatherhead East Asian Institute and Data Science Institute at Columbia University.

Competing interests. None.

References

- Andrews, Kenneth T., and Neal Caren.** 2010. "Making the news: movement organizations, media attention, and the public agenda." *American Sociological Review* 75 (6), 841–866.
- Angrist, Joshua D., and Jörn-Steffen Pischke.** 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- Barranco, José, and Dominique Wisler.** 1999. "Validity and systematicity of newspaper data in event analysis." *European Sociological Review* 15 (3), 301–322.
- Cai, Yongshun.** 2010. *Collective Resistance in China: Why Popular Protests Succeed or Fail*. Stanford, CA: Stanford University Press.
- Caren, Neal, Kenneth T. Andrews and Todd Lu.** 2020. "Contemporary social movements in a hybrid media environment." *Annual Review of Sociology* 46 (1), 443–465.
- Chen, Chih-Jou Jay.** 2020. "A protest society evaluated: popular protests in China, 2000–2019." *Mobilization: An International Quarterly* 25 (SI), 641–660.
- Chen, Xi.** 2009. "The power of 'troublemaking': protest tactics and their efficacy in China." *Comparative Politics* 41 (4), 451–471.
- Distelhorst, Greg, and Yue Hou.** 2017. "Constituency service under nondemocratic rule: evidence from China." *Journal of Politics* 79 (3), 1024–40.
- Earl, Jennifer.** 2015. "The future of social movement organizations: the waning dominance of SMOs online." *American Behavioral Scientist* 59 (1), 35–52.
- Earl, Jennifer, Andrew Martin, John D. McCarthy and Sarah A. Soule.** 2004. "The use of newspaper data in the study of collective action." *Annual Review of Sociology* 30, 65–80.
- Eberl, Jakob-Moritz, Hajo G. Boomgaarden and Markus Wagner.** 2017. "One bias fits all? Three types of media bias and their effects on party preferences." *Communication Research* 44 (8), 1125–48.
- Elfstrom, Manfred, and Sarosh Kuruvilla.** 2014. "The changing nature of labor unrest in China." *ILR Review* 67 (2), 453–480.
- Fang, Kecheng.** 2022. "What is *zimeiti*? The commercial logic of content provision on China's social media platforms." *Chinese Journal of Communication* 15 (1), 75–94.
- Fong, Christian, and Matthew Tyler.** 2020. "Machine learning predictions as regression covariates." *Political Analysis* 29 (4), 1–18.
- Francisco, Ronald A.** 2005. "The dictator's dilemma." In Christian Davenport, Hank Johnston and Carol Mueller (eds.), *Repression and Mobilization*. Minneapolis: University of Minnesota Press, 58–82.
- Goebel, Christian, and Lynette H. Ong.** 2012. "Social unrest in China." SSRN Scholarly Paper ID 2173073. Rochester, NY: Social Science Research Network.
- Hale, Henry E.** 2013. "Regime change cascades: what we have learned from the 1848 revolutions to the 2011 Arab uprisings." *Annual Review of Political Science* 16 (1), 331–353.
- Hong, Sounman.** 2012. "Online news on Twitter: newspapers' social media adoption and their online readership." *Information Economics and Policy* 24 (1), 69–74.
- Hug, Simon, and Dominique Wisler.** 1998. "Correcting for selection bias in social movement research." *Mobilization: An International Quarterly* 3 (2), 141–161.
- Hutter, Swen.** 2014. "Protest event analysis and its offspring." In Donatella Della Porta (ed.), *Methodological Practices in Social Movement Research*. Oxford: Oxford University Press, 335–367.

- Iyengar, Shanto.** 1991. *Is Anyone Responsible?: How Television Frames Political Issues*. American Politics and Political Economy Series. Chicago: University of Chicago Press.
- Ju, Alice, Sun Ho Jeong and Hsiang Iris Chyi.** 2014. "Will social media save newspapers?" *Journalism Practice* 8 (1), 1–17.
- Koopmans, Ruud.** 2004. "Movements and media: selection processes and evolutionary dynamics in the public sphere." *Theory and Society* 33 (3–4): 367–391.
- Lee, Ching Kwan, and Yonghong Zhang.** 2013. "The power of instability: unraveling the microfoundations of bargained authoritarianism in China." *American Journal of Sociology* 118 (6), 1475–1508.
- Lei, Ya-Wen.** 2016. "Freeing the press: how field environment explains critical news reporting in China." *American Journal of Sociology* 122 (1), 1–48.
- Lorentzen, Peter.** 2017. "Designing contentious politics in post-1989 China." *Modern China* 43 (5), 459–493.
- McCarthy, John D., Clark McPhail and Jackie Smith.** 1996. "Images of protest: dimensions of selection bias in media coverage of Washington demonstrations, 1982 and 1991." *American Sociological Review* 61 (3), 478–499.
- Oliver, Pamela E., and Gregory M. Maney.** 2000. "Political processes and local newspaper coverage of protest events: from selection bias to triadic interactions." *American Journal of Sociology* 106 (2), 463–505.
- Ong, Lynette.** 2015. "Reports of social unrest: basic characteristics, trends and patterns, 2003–12." In David S. G. Goodman (ed.), *Handbook of Research on Politics in China*. Cheltenham, UK: Edward Elgar, 345–360.
- Qin, Bei, David Strömberg and Yanhui Wu.** 2018. "Media bias in China." *American Economic Review* 108 (9), 2442–76.
- Repnikova, Maria, and Kecheng Fang.** 2018. "Authoritarian participatory persuasion 2.0: netizens as thought work collaborators in China." *Journal of Contemporary China* 27 (113), 763–779.
- Roberts, Margaret E.** 2018. *Censored: Distraction and Diversion inside China's Great Firewall*. Princeton, NJ: Princeton University Press.
- Shao, Dongke.** 2017. "The construction and application of mass incidents database in China." *China Public Administration* 381 (3), 126–130.
- Smith, Jackie, John D. McCarthy, Clark McPhail and Boguslaw Augustyn.** 2001. "From protest to agenda building: description bias in media coverage of protest events in Washington, D.C." *Social Forces* 79 (4), 1397–1423.
- State Council of the PRC.** 2018. "Guanyu tuijin zhengwu xin meiti jiankang youxu fazhan de yijian" (Opinions on promoting the healthy and orderly development of new media in government affairs). www.gov.cn/zhengce/content/2018-12/27/content_5352666.htm. Accessed 9 June 2023.
- Stockmann, Daniela, and Mary E. Gallagher.** 2011. "Remote control: how the media sustain authoritarian rule in China." *Comparative Political Studies* 44 (4): 436–467.
- Stockmann, Daniela, and Ting Luo.** 2017. "Which social media facilitate online public opinion in China?" *Problems of Post-Communism* 64 (3–4): 189–202.
- Su, Zheng, and Tianguang Meng.** 2016. "Selective responsiveness: online public demands and government responsiveness in authoritarian China." *Social Science Research* 59, 52–67.
- Tang, Wenfang.** 2016. *Populist Authoritarianism: Chinese Political Culture and Regime Sustainability*. Oxford: Oxford University Press.
- Teets, Jessica C.** 2013. "Let many civil societies bloom: the rise of consultative authoritarianism in China." *The China Quarterly* 213, 19–38.
- Truex, Rory.** 2017. "Consultative authoritarianism and its limits." *Comparative Political Studies* 50 (3), 329–361.
- Wang, Yuhua, and Carl Minzner.** 2015. "The rise of the Chinese security state." *The China Quarterly* 222, 339–359.
- Wolfsfeld, Gadi, Elad Segev and Tamir Sheafer.** 2013. "Social media and the Arab Spring: politics comes first." *International Journal of Press/Politics* 18 (2), 115–137.
- Zhang, Han, and Jennifer Pan.** 2019. "CASM: A deep-learning approach for identifying collective action events with text and image data from social media." *Sociological Methodology* 49 (1), 1–57.

Han ZHANG is an assistant professor in the Division of Social Science at the Hong Kong University of Science and Technology. His research interests include computational social science, social movements, and political sociology. His research won best paper awards from the Section on Collective Behavior and Social Movements of the American Sociological Association and the Computational Methods Division of the International Communication Association.

Yao LU is a professor of Sociology and faculty affiliate of the Weatherhead East Asian Institute and Data Science Institute at Columbia University. Her research is at the intersection of inequality, demography and politics. Her recent work examines the influence of demographic forces and inequality on political processes using surveys, experiments and computational data.

Rui BAI is a graduate student in the Data Science Institute at Columbia University. Her research focuses on machine learning and natural language processing. She also works on AI for social good.

Cite this article: Zhang H, Lu Y, Bai R (2024). Selection and Description Bias in Protest Reporting by Government and News Media on Weibo. *The China Quarterly* 257, 75–99. <https://doi.org/10.1017/S0305741023001042>