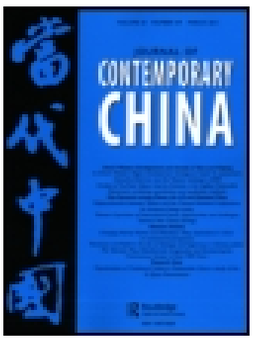

Journal of Contemporary China

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cjcc20

The Political Logic of Protest Repression in China

Christian Göbel

To cite this article: Christian Göbel (2020): The Political Logic of Protest Repression in China, Journal of Contemporary China, DOI: 10.1080/10670564.2020.1790897

To link to this article: https://doi.org/10.1080/10670564.2020.1790897

JOURNAL OF CONTEMPORARY CHINA
https://doi.org/10.1080/10670564.2020.1790897

The Political Logic of Protest Repression in China

Christian Göbel

University of Vienna, Austria

ABSTRACT

Why do China's authorities repress some protests, but not others? By how much do crowd size, violent tactics and protest location increase the likelihood of repression? Based on a newly available dataset of more than 70,000 protest events collected from social media, this article tests three competing explanations of protest repression in China. It finds that repression is closely correlated both with the cost of concessions for local governments and protest intensity. A small-scale and peaceful labor protest in an urban locality very seldom encounters repression, but rural riots against land grabs, evictions or environmental pollution are nearly certain to experience state-sanctioned violence or arrests even if the number of participants is low.

Introduction

On 28 November 2019, a riot broke out in Wenlou Township, one day after citizens had learned that a planned ecological park would include a crematorium.[1] The government reacted swiftly by deploying riot police, which roughed up and arrested protesters. This incident came only four months after thousands of citizens in Yangluo County demonstrated against the planned construction of a garbage incinerator plant, which also was met with fierce repression.[2] While these and other examples might suggest that government repression of popular protest in China is the norm, most protests do not meet with suppression. The Chinese authorities often tolerate social unrest because it can help secure people's rights against a fragmented, patrimonial and inefficient bureaucracy.[3]

Why do the Chinese authorities repress some protest events, but not others? Existing research provides three explanations. The first sees physical coercion as a last resort: according to this explanation, the authorities are generally responsive to protests and only apply repression when all other measures have failed.[4] The second regards repression as a strategic deterrent to raise the opportunity costs of engaging in social unrest.[5] The third regards repression as a function of cost/benefit calculations by local officials, and the 'forcefulness' of a protest: repression will be used when it is more beneficial for them than making concessions, or when a protest lacks

CONTACT Christian Göbel christian.goebel@univie.ac.at

[1]Lily Kuo, 'Chinese Riot Police Fire Teargas and Beat Up Protesters in Guangdong Province', *The Guardian*, November 30, 2019, accessed January 27, 2020, https://www.theguardian.com/world/2019/nov/30/chinese-riot-police-fire-teargas-and-beat-up-protesters-in-guangdong-province

[2]Pinghui Zhuang, 'Thousands Protest in Central China Over Waste Incineration Plant', *South China Morning Post*, July 5, 2019, accessed January 27, 2020, https://www.scmp.com/news/china/society/article/3017386/thousands-protest-central-china-over-waste-incineration-plant

[3]Xi Chen, *Social Protest and Contentious Authoritarianism in China* (Cambridge: Cambridge University Press, 2012); Yao Li, 'A Zero-Sum Game? Repression and protest in China', *Government and Opposition* 54(2) (2017), pp. 1–27; Yao Li, *Playing by the Informal Rules: Why the Chinese Regime Remains Stable Despite Rising Protests* (Cambridge: Cambridge University Press, 2018).

[4]Ibid.

[5]Peter Lorentzen, 'Designing contentious politics in post-1989 China', *Modern China* 43(5) (2017): pp. 459–93.

'forcefulness.'[6] 'Forcefulness' refers to factors such as protest size, protester violence, repression and media coverage and is thus similar to the concept of protest 'intensity', which will be used in this article.

The three explanations are only partly compatible—as opposed to the first and third explanations, explanation two claims that repression is essentially random. Explanations 1 and 3 are partly compatible in that the specific features of a protest, i.e. its intensity, can provide officials with a convenient excuse to crack down on protests where the cost of concessions is high.

An important reason why so little is known about the determinants of repression in China is a dearth of structured data—detailed protest statistics are not available, domestic news media are heavily censored, and reports on Chinese protests in international media are rare and likely biased.[7] Most existing studies are therefore informed either by single cases, and only a few draw on datasets that contain a few hundred observations at most. More fine-grained and comprehensive data is needed to account for the various factors that might cause the repression of protesters. Such data has recently become available in the form of a website containing more than 70,000 protest events that occurred in all over China between July 2013 and June 2016. The events were hand-collected by Lu Yuyu and Li Tingyu with the aim of creating a platform that would at once 'draw attention' to protesters' struggles and facilitate learning about what makes a protest effective.[8]

The author downloaded all protest events on the website and classified each event according to its underlying issue and intensity. Inferential statistics are used to test if repression is the function of protest intensity (explanation 1, explanation 3) or of the financial costs of government concessions (explanation 3). If none of these factors is found to have an impact, i.e. if they are randomly distributed, this is regarded as support for explanation 2.

The results provide support for explanation 3: repression is conditional on how much it costs the government to make a concession. Controlling for protest locality, time and intensity, issues where protesters demand compensation from the government are far more likely to encounter repression than protests directed against private actors. Financial compensation for land grabs, evictions or medical mistakes would have to be made by the government but making up for the financial losses suffered by defrauded home-owners and wages owed to migrant workers is the responsibility of private companies. If protests aim at securing financial compensation by the government, the likelihood of repression is well above 50 percent. If the target is private actors, it is below 20 percent. They also provide some support for explanation 1: violent events are more likely to face repression than non-violent events. Hence, repression is non-random, which seems to speak against explanation 2. However, it is worth noting that even though repression is more likely for 'costly' issues, not all these events will be repressed by the government. This article explains why land protests are more likely to be repressed than labor protests, but it cannot explain why some land protests met with repression, but not others. Here, explanation 2 might be of value.

Determinants of Repression

Why governments use repression against some protests, but not others are one of the big unresolved questions of political science.[9] Existing studies show that anti-regime protests are especially

[6]Yongshun Cai, *Collective Resistance in China: Why Popular Protests Succeed or Fail* (Stanford: Stanford University Press, 2010), p. 45.

[7]Christian Göbel and H Christoph Steinhardt, 'Better coverage, less bias: using social media to measure protest in authoritarian regimes', *ResearchGate* (2019), accessed May 27, 2020, doi: 10.13140/RG.2.2.32856.75523/1.

[8]China Labour Bulletin, 'Lu Yuyu and Li Tingyu, the Activists Who Put Non-News in the News', August 18, 2017, accessed January 27, 2020, https://clb.org.hk/content/lu-yuyu-and-li-tingyu-activists-who-put-non-news-news

[9]Christian Davenport, 'State repression and political order', *Annu. Rev. Polit. Sci.* 10 (2007), pp. 1–23; Christian Davenport and Molly Inman, 'The state of state repression research since the 1990s', *Terrorism and Political Violence* 24(4) (2012), pp. 619–34.

likely to incur government repression.[10] Beyond that, there is little agreement on the factors that cause political leaders to order the arrest or use of violence against protesters.[11]

The idiosyncratic nature of each protest is a possible reason for the dearth of robust findings on the determinants of repression. Repression is inherently risky, because crackdowns can easily spiral out of control.[12] There is a risk that protesters fight back, and that bystanders take the side of the protesters. Hence, repression can transform a peaceful protest into a violent uprising.[13] The occurrence or non-occurrence of repression is thus generally understood as the outcome of cost-benefit calculations by the authorities.[14] However, the cost of repression is highly contingent and therefore hard to predict even for political leaders who dispose of far more information than political scientists. If, as these rational choice models assume, the decision to repress or not to repress is made on a case by case basis, formulating models that capture this complexity, yet are applicable to a wide range of protests, is indeed challenging.

This is true not only for the cross-national level, but also for the variation of repression within China. There is a rich tradition of protest research within the field of China studies,[15] but this research has almost exclusively focused on the actions of the protesters. Under what conditions popular protest is met by government repression remains understudied.[16] Most of the little research that does exist shares the basic assumptions of the cross-national theories discussed above. There is no doubt that perceived anti-regime activities, which form a minuscule tiny minority of all protests, will be forcefully subdued[17] but that otherwise, repression and non-repression are contingent on individual-case cost-benefit calculations. Klein and Regan make a useful distinction between concessions costs and disruption costs.[18] Concession costs are high when protesters express 'maximalist' demands such as 'high-level resignations or changes in political representation', use violence, or make repeated demands.[19] Disruption costs are high when protest activities disrupt business, endure for a long time and attract a large crowd.[20] Analyzing 13,015 medium and large-scale protests, the authors find that high concession costs invite repression, and high disruption costs accommodation if concession costs are not also high. If concession- and disruption costs are both low, protests tend to be disregarded.

[10]Nathan Danneman and Emily Hencken Ritter, 'Contagious rebellion and preemptive repression', *Journal of Conflict Resolution* 58(2) (2014), pp. 254–79, Christian Davenport, 'State Repression and Political Order', *Annu. Rev. Polit. Sci*. 10 (2007), p. 7; for China, see Rory Truex, 'Focal points, dissident calendars, and preemptive repression', *Journal of Conflict Resolution* 63(4) (2019), pp. 1032–52.

[11]Christian Davenport, 'State repression and political order', *Annu. Rev. Polit. Sci*. 10 (2007), p. 8.

[12]Jennifer Earl and Sarah A. Soule, 'The impacts of repression: The effect of police presence and action on subsequent protest rates', *Research in Social Movements, Conflicts and Change* 30 (2010), pp. 75–113; Paul D. Almeida, 'Opportunity organizations and threat-induced contention: Protest waves in authoritarian settings', *American Journal of Sociology* 109(2) (2003), pp. 345–400.

[13]The popular demonstrations that preceded the 1989 Tiananmen massacre are one example, the 1988 strikes in Poland that were later joined by members of other social strata another.

[14]Christian Davenport, 'State repression and political order', *Annu. Rev. Polit. Sci*. 10(2007), p. 17.

[15]for book-length treatises on the subject besides those already cited above, see for example Diana Fu, *Mobilizing Without the Masses: Control and Contention in China* (Cambridge: Cambridge University Press, 2017); Kevin J. O'Brien, *Popular Protest in China* (Harvard University Press, 2009); Kevin J. O'Brien and Lianjiang Li, *Rightful Resistance in Rural China* (Cambridge University Press, 2006).

[16]notable exceptions that address this question are Manfred Elfstrom, 'Two steps forward, one step back: Chinese state reactions to labour unrest', *The China Quarterly* 240 (2019), pp. 1–25; Yao Li, 'A zero-sum game? Repression and protest in China', *Government and Opposition* 54(2) (2017), pp. 1–27; Yanqi Tong and Shaohua Lei, 'Large-scale mass incidents and government responses in China', *International Journal of China Studies* 1(2) (2010), pp. 487–508; Chih-Jou Chen, 'Growing social unrest and emergent protest groups in China', in *Rise of China: Beijing's Strategies and Implications for the Asia-Pacific*, ed. Chang-yi Lin and Xinhuang Xiao (London:Routledge, 2009), pp. 87–105; Yongshun Cai, 'Local governments and the suppression of popular resistance in China', *The China Quarterly* 193 (2008), pp. 24–42

[17]cases in point are the suppression of the Tian'anmen protests in June 1989 and the deteriorating human rights situation in Xinjiang.

[18]Graig R. Klein and Patrick M. Regan, 'Dynamics of political protests', *International Organization* 72 (2018), pp. 485–521

[19]Ibid, p. 492.

[20]Ibid, p. 493.

The main difference between transnational studies of protest repression and their China studies counterparts is that the former regards protest as a liability for a regime, whereas the latter has produced insights on the systemic benefits of popular contention. Cai Yongshun's approach probably comes closest to the former perspective. He models protest outcomes as a function of two factors: the cost of concessions for the authorities and the 'forcefulness' of protesters' actions. Cai distinguishes between two types of 'costs'. One is political, the other monetary costs. If concessions are likely to encourage more resistance[21] or would require punishing influential leaders, the political costs are high. The monetary costs of concession are high if leaders have to tap government finances to meet protesters' demands, or if they have a negative effect on revenue such as land transfer fees (in Regan and Klein's typology, these constitute 'low threat' demands).[22] Cai also holds that if a protest has more than 500 participants, involves casualties, or receives media coverage,[23] the authorities are likely to make concessions. This dimension is not directly relevant for repression, because repression frequently precedes concessions. This was the case, for example, in the protest described in the introduction. The authorities forcefully repressed the anti-crematorium riot in Wenlou Township, but later suspended the planned construction.[24]

Others, however, stress the potential benefits of protests that do not challenge the regime. This line or argument rests on the assumption that there exists a constellation between state and social forces in China, which is different from most other authoritarian countries. Xi Chen argues that popular protests are not necessarily a threat to, but can instead help stabilize China's one-party authoritarian regime, an argument also made by Yao Li.[25] According to this line of argument, the CCP opens limited space for protests so people can protect their rights and interests. Protests provide higher levels with insights into severe grievances that lower level governments would try to hush up. Repression, according to Chen, is 'infrequent' and, if it occurs, 'restrained.'[26]

Peter Lorentzen's contribution agrees with Chen's in that protests can serve the regime but provides further insights as to when the authorities are likely to use repression.[27] According to Chen, repression is a means of last resort, but according to Lorentzen, it is strategic. Repression imposes a cost on protesting. If the risk of getting arrested or beaten up is too low, citizens will take to the street over trifle matters. If the risk is too high, only those with nothing to lose will protest. Hence, it is in the interest of the authorities to dose repression in such a way that they obtain the information they need without encouraging protests that are of no benefit to them.[28] This resonates with the contentious politics literature, which has established that state repression influences the emergence of future protests.

The three models have different implications for when repression is likely to occur. Xi Chen, for example, does not make any predictions as to when the authorities see no other way but to use violence against protesters. Lorentzen holds that protesters should not be able to predict repression. This means that repression should occur randomly. Repression should be applied in such a way that it imposes a barrier high enough for protesters to not take to the streets without a good reason, but not so high as to make them believe that protest is essentially futile.[29] However, Lorentzen makes no proposition as to where this threshold might lie in practice. Finally, Cai's model predicts that

[21]Yongshun Cai, *Collective Resistance in China: Why Popular Protests Succeed or Fail* (Stanford: Stanford University Press, 2010), p. 17.
[22]Ibid., Chapter 3, Graig R. Klein and Patrick M. Regan, 'Dynamics of political protests', *International Organization* 72 (2018), p. 500.
[23]Yongshun Cai, *Collective Resistance in China: Why Popular Protests Succeed or Fail* (Stanford: Stanford University Press, 2010), p. 44.
[24]Inkstone, 'Chinese city backs down after protests (No, it's not Hong Kong)', December 2019, accessed January 27, 2020, https://www.inkstonenews.com/politics/chinese-city-wenlou-suspends-crematorium-plan-after-protests/article/3040183
[25]Yao Li, 'A zero-sum game? Repression and protest in China', *Government and Opposition* 54(2) (2017), pp. 1–27.
[26]Xi Chen, *Social Protest and Contentious Authoritarianism in China* (Cambridge: Cambridge University Press, 2012).
[27]Peter Lorentzen, 'Designing contentious politics in post-1989 China', *Modern China* 43(5) (2017): pp. 459–93.
[28]Ibid.
[29]Ibid.

repression will occur when the cost of making concessions is perceived higher than the cost of beating up or arresting protesters, or when a protest is not 'forceful', i.e. intense, enough.

Operationalizing Repression, Protest Intensity and Cost of Concessions

This study investigates the impact of two factors on the likelihood of repression: protest intensity, and the costs local governments would have to bear if they conceded to the demands of protesters. Following Davenport, state repression is defined as 'the actual or threatened use of physical sanctions against an individual or organization, within the territorial jurisdiction of the state.'[30] As most social media representations of protest events do not mention threats made against protesters, this source cannot be used to reliably measure the 'threatened use of physical sanctions'. For this reason, the definition of repression is narrowed down to entail only the 'actual use of physical sanctions'. Physical sanctions come in two forms: violence and arrests, so a protest event will be coded as 'repressed' when at least one protester was arrested or assaulted by an agent of the state. Agents of the state include police officers, riot police, urban law enforcement officials, but also paramilitary forces and thugs hired by local officials.[31]

The independent variables are protest intensity and the cost of concessions. Protest intensity relates to the size and duration of a protest, if protesters use violence, and if the authorities use repression. The larger, sustained and more violent a protest, the more intense it is.[32] Unfortunately, there exists no information on the duration of most protests in the dataset. Social media posts on intense protests are likely to be reposted often, and therefore deleted quickly.[33] There is, however, information on the size of a protest, and whether protesters' behavior is peaceful or violent. The articles use protest intensity instead of Cai's concept of 'forcefulness', because the former is a more nuanced instrument. Most protests in China are not forceful in the sense that they draw a large protest crowd, claim casualties or are reported in national or international media. Out of 2,500 protest events in the three megacities Guangzhou, Shanghai and Chongqing, only 80 had more than 500 participants, and less than 40 percent of those were reported in national or international news media.[34] Applying Cai's definition, this means that only a small fraction of all protests can be considered forceful. If forcefulness influenced repression independent of the cost of concessions, its effect would be very small. In contrast, protest intensity can be measured for any protest event, no matter how small or violent.

As for the cost of concessions, this article uses a much simpler approach to measurement than the one proposed by Cai. Cai simply uses a binary variable to measure if costs are 'high' or 'low',[35] because there is not enough information to calculate the cost of concessions for each protest in monetary terms. Examining each individual protest event is a high price to pay for a measure as crude as a binary variable, and not feasible for a dataset that consists of more than 70,000 events. Instead, this article proposes to use protest issue as a proxy for the cost of concessions. This is possible because the two are closely related. Land protests, for example, are associated with high cost of concessions, because the compensation for farmers needs to be paid by the government. The same is true for evictions, medical issues, and rural 'Not in My Backyard' (NIMBY) environmental

[30]Christian Davenport, 'State repression and political order', *Annu. Rev. Polit. Sci.* 10(2007), pp. 1–23.
[31]Lynette H. Ong, '"Thugs-for-Hire": Subcontracting of state coercion and state capacity in China', *Perspectives on Politics* 16(3) (2018), pp. 680–95.
[32]David Snyder and William R. Kelly, 'Conflict intensity, media sensitivity and the validity of newspaper data', *American Sociological Review* 42(1) (1977), pp. 105–23; Daniel J. Myers and Beth Schaefer Caniglia, 'All the rioting that's fit to print: selection effects in national newspaper coverage of civil disorders, 1968–1969', *American Sociological Review* 69(4) (2004), pp. 519–43.
[33]Gary King, Jennifer Pan, and Margaret E. Roberts, 'How censorship in China allows government criticism but silences collective expression', *American Political Science Review* 107(2) (2013), pp. 326–43.
[34]Christian Göbel and H Christoph Steinhardt, 'Better coverage, less bias: Using social media to measure protest in authoritarian regimes', *ResearchGate* (2019), accessed May 27, 2020, doi: 10.13140/RG.2.2.32856.75523/1.
[35]Yongshun Cai, *Collective Resistance in China: Why Popular Protests Succeed or Fail* (Stanford: Stanford University Press, 2010), p. 45.

protests. In contrast, the cost of concessions for labor protests is low, because the government can order factory owners to pay higher wages or respect labor contracts. They might need to make concessions to the factory owners, for example by giving them access to subsidized land or improving transport infrastructure, but the protests that might result from land expropriation can once more be repressed at low cost.

There are two distinct advantages to this approach: first, it is scalable. Issue types can be classified with the help of machine learning algorithms, which facilitates the processing of large datasets. Second, measuring the effect of issue type on repression provides readers with information as to which types of protesters are most likely to face government repression.

Instead of using newspapers, as Cai did, this article draws on protest event data sourced from social media. Social media has the distinct advantage that protest coverage is far more comprehensive and selection bias is less pronounced than in traditional media.[36] However, protest events in remote regions where cellphone- and Internet penetration is low are likely to be underrepresented,[37] and events that cause 'volume bursts' on social media are likely to be censored.[38]

The Wickedonna Dataset

A dataset of social media posts covering 74,425 protest events between July 2013 and June 2016 underlies the analysis. The data was collected by Lu Yuyu and Li Tingyu, two activists who named their project 'Wickedonna'. Lu and Li searched and selected every entry by hand. They ensured the veracity of each event by collecting photos from different sources and, in some cases, contacting participants on social media. The results were uploaded to a publicly accessible blog.[39] There are no more entries after 13 June 2016, shortly before Lu and Li were arrested on charges of 'picking quarrels.'[40] The Wickedonna dataset provides the most complete coverage of Chinese protest events between July 2013 and June 2016, outperforming all other sources by a wide margin. Its bias is estimated to be smaller than that of traditional media and dissident websites and is manifested mainly in the underrepresentation of protests in areas with low mobile phone density.[41] This is the case in underdeveloped rural regions. Protests in minority regions are most likely also underrepresented, not only because they tend to be underdeveloped, but because information control there is even stronger than in the rest of China. Finally, there are no anti-regime protests in the database. If they did take place, social media posts on them would have been censored quickly.

Lu Yuyu and Li Tingyu annotated 52,629 of 74,425 events on the website with keywords, which were regrouped into 11 distinct issues (labor, realestate, fraud, landgrab, eviction, medical issues, education, transport, corruption, police brutality, pollution) (Appendix A). A Support Vector Machine (SVM) was employed to classify unlabeled events. The label 'police brutality' was dropped from the 577 events that were labelled such. Almost all of these events include keywords that are related to repression, which makes it is impossible for the algorithm to detect if repression followed a protest or if a protest followed repression. Most of these events are related to hawkers or retailers that resisted forced eviction or rent hikes, experienced repression, and then vented their anger online. The label was dropped, but the events remain in the sample, which allows the SVM to classify them according to the underlying issue (e.g. eviction, real-estate …). On average, 93 percent of the labels assigned

[36]Christian Göbel and H Christoph Steinhardt, 'Better coverage, less bias: Using social media to measure protest in authoritarian regimes', *ResearchGate* (2019), accessed May 27, 2020, doi: 10.13140/RG.2.2.32856.75523/1.

[37]Nils B. Weidmann, 'A closer look at reporting bias in conflict event data', *American Journal of Political Science* 60(1) (2016), pp. 206–18.

[38]Gary King, Jennifer Pan, and Margaret E. Roberts, 'How censorship in China allows government criticism but silences collective expression', *American Political Science Review* 107(2) (2013), pp. 326–43.

[39]see the website at http://newsworthknowingcn.blogspot.com

[40]China Labour Bulletin, 'Lu Yuyu and Li Tingyu, the activists who put non news in the news', August 18, 2017, accessed January 27, 2020, https://clb.org.hk/content/lu-yuyu-and-li-tingyu-activists-who-put-non-news-news

[41]Christian Göbel and H Christoph Steinhardt, 'Better coverage, less bias: Using social media to measure protest in authoritarian regimes', *ResearchGate* (2019), accessed May 27, 2020, doi: 10.13140/RG.2.2.32856.75523/1.

by the SVM are correct (precision: 93 percent, recall: 93 percent), with the best performing category (transport) having a precision score of 96 percent, the worst (rural land) of 87 percent.

Estimating the size of a protest is very challenging. In very small protests, participants can still be counted, but counts become more imprecise as protest sizes increase in number. In addition, protesters tend to over-report, and the government under-report protest sizes. To tackle these challenges, researchers have experimented with alternative measures such as cell phone location data, the number of Twitter accounts that contain the hashtag of a protest event, and the number of faces contained in protest images shared on social media.[42] Sobolev et. al. find that all alternative measures provide reliable estimates, even though they slightly over-estimate the size of small and under-estimate the size of large protests.

This article uses an artificial neural network that provides estimates based on density maps. Simply put, images with known numbers of participants are transformed into density maps, which are used to predict crowd sizes in new data. A ResNet-50 convolutional neural network was trained on the ShanghaiTech Part B dataset, which contains 716 images of busy streets and squares in metropolitan Shanghai.[43] For each event, the crowd size estimates of all images were added up. As the number of images of an event on the Wickedonna website increases with event size, estimates are probably biased upwards for larger events. To mitigate this effect, the natural logarithm of each size estimate was used in the analysis. Those events that do not contain images with persons in them (e.g. banners on house walls) are excluded, reducing the sample to 71,035 events.

Machine learning classification proved impractical for classifying protests according to the presence or absence of government repression. Given the close association between police and repression, the algorithms frequently misclassified a mere presence of police as repression. To address this challenge, this study employs a keyword-based approach for classification.

To capture as many relevant and exclude as many irrelevant terms as possible, the following procedure is applied: first, while coding for protest issue, the author noted down synonyms for violence and arrests. This yielded the most frequent terms used for both phenomena. Then, a -Word2Vec[44] model was trained on the full, tokenized corpus of social media protest posts. Word2Vec is a shallow, two-layer neural network that produces word embeddings, i.e. measures on the semantic similarity between the words in a corpus. The model was trained with the Word2Vec Python library *gensim*.[45] The model was used to yield the 50 most similar terms for each relevant word identified in the previous step, and then the five most similar terms for each of the 50 results. Finally, duplicate entries were deleted, and the resulting list of unique terms related to each phenomenon were manually inspected to ensure they are related to the phenomenon of interest. A dummy variable was created to measure each phenomenon. It scores 1 if an entry contains any of the terms in the corresponding list and 0 if it does not. The dictionaries are listed in Appendix B.

Protest Issues

Before moving on to the analysis, this section briefly introduces the main issues that drive China's citizens to the streets. Based on the introduction of the three main theories in Chapter 2, which effect each issue should have on repression is also outlined below. Table 1 provides the summary statistics for the dependent, independent and control variables used in the statistical analysis in the section below.

[42]Anton Sobolev et al., 'News and Geolocated Social Media Accurately Measure Protest Size', accessed May 27, 2020, https://www.anderson.ucla.edu/faculty_pages/keith.chen/papers/WP_MeasuringProtestSize.pdf

[43]Yingying Zhang et al., 'Single-Image Crowd Counting via Multi-Column Convolutional Neural Network', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 589–97.

[44]Yoav Goldberg and Omer Levy, 'Word2vec Explained: Deriving Mikolov et Al.'s Negative-Sampling Word-Embedding Method', *ArXiv Preprint*, 2014, accessed on May 27, 2020, https://arxiv.org/abs/1402.3722

[45]Gensim, 'models.word2vec—Word2vec embeddings', accessed January 27, 2020, https://radimrehurek.com/gensim/models/word2vec.html

Table 1. Summary statistics

	count	mean	stderr	min	Max
Issues					
Labor	74416	0.429	0.495	0.0	1.0
Realestate	74416	0.209	0.407	0.0	1.0
rural land	74416	0.101	0.301	0.0	1.0
Medical	74416	0.064	0.244	0.0	1.0
Eviction	74416	0.064	0.244	0.0	1.0
Fraud	74416	0.045	0.207	0.0	1.0
Education	74416	0.026	0.159	0.0	1.0
Pollution	74416	0.023	0.150	0.0	1.0
Repression					
Repression	74416	0.250	0.433	0.0	1.0
Intensity					
demonstrate	74416	0.208	0.406	0.0	1.0
Riot	74416	0.137	0.344	0.0	1.0
Petition	74416	0.165	0.371	0.0	1.0
Strike	74416	0.013	0.112	0.0	1.0
Size	71278	232.45	1183.43	0.0	142284
Controls					
Village	74416	0.279	0.449	0.0	1.0

Eight protest issues make up most of all measured protests in China.

1) Labor protests are by far the most common motivator for a protest. Nearly 43 percent of all protests are related to unpaid wages, arrears and other labor issues. Migrant workers constitute the main group of protesters,[46] protests by taxi drivers, bus drivers and teachers are less common. Labor related grievances predominately occur in privately owned enterprises, who would have to bear the cost of concessions.[47] Accordingly, the cost of concessions is low for the government, so repression levels should also be low.

2) Real-estate related grievances motivate almost 21 percent of all protests. This category includes conflicts between owners and developers as well as between owners and property management companies.[48] The former pertains to apartments that are not delivered on time or in the quality agreed on in the contract, but also to cases where developers default on pre-contracts. In such cases, prospective homeowners might lose all their investments. Property management companies are the object of homeowners' ire because they are often accused of overcharging residents or providing sub-standard services. Again, concessions would have to be made by private companies and not the government, so the level of repression against real estate protest should also be low.

3) Land grabs account for 10 percent of all protests. Land grabs are a predominantly rural phenomenon and occur when land is requisitioned either for real estate development or commercial purposes.[49] Rural real estate development might serve to provide affordable housing to residents of nearby cities or to concentrate rural residents in designated residential areas to free up farmland.

[46]Mary E. Gallagher, 'China's workers movement & the end of the rapid-growth era', *Daedalus* 143(2) (2014), pp. 81–95; Ching Kwan Lee, 'Pathways of labor insurgency', *Chinese Society: Change, Conflict and Resistance* (41) (2000), p. 61; Anita Chan, 'Recent trends in Chinese labour issues: Signs of change', *China Perspectives* (57) (2005), accessed May 27, 2020, https://doi.org/10.4000/chinaperspectives.1115

[47]as of May 2019, only 100 out of 1,338 labor protests recorded by the China Labour Bulletin were directed against state owned enterprises (China Labour Bulletin, 'Strike Map', accessed January 27 2020, https://maps.clb.org.hk/strikes/en)

[48]Luigi Tomba, *The Government Next Door: Neighborhood Politics in Urban China* (Ithaca NY: Cornell University Press, 2014); Ying Wu and Junhua Chen, 'The constructive significance of homeowners' rightful protest in China', in *Neighbourhood Governance in Urban China*, ed. Ngai-Ming Yip (Cheltenham:Edward Elgar Publishing, 2014), p. 167; Yongshun Cai, 'China's Moderate Middle Class: The Case of Homeowners' Resistance', *Asian Survey* 45(5) (2005), pp. 777–99.

[49]Sally Sargeson, 'Violence as development: land expropriation and China's urbanization', *Journal of Peasant Studies* 40(6) (2013), pp. 1063–85; You-tien Hsing, *The Great Urban Transformation: Politics of Land and Property in China* (Oxford: Oxford University Press, 2012); Kathy Le Mons Walker, 'From covert to overt: Everyday peasant politics in China and the implications for transnational agrarian movements', *Journal of Agrarian Change* 8(2–3) (2008), pp. 462–88; Xiaolin Guo, 'Land expropriation and rural conflicts in China', *The China Quarterly* 166(2001), pp. 422–39.

Rural officials also requisition land to attract factory owners, who invest in the countryside because of low factor costs or the availability of natural resources. Garbage incinerators and chemical plants frequently constructed in the countryside in response to increasingly strict pollution regulations. Not-in-my-backyard protests and the negative effect of polluting industries on urban property values also are additional push-factors to locate such projects in the countryside.[50] In the late 2000 s, land sales made up around 2/3 of the revenue of China's city governments.[51] Given the importance of land revenue for local finances, the cost of concessions is high, so the level of repression should be high, too.

4) Six percent of all protests are against evictions. Like the requisition of rural land, forced evictions are a means to generate finances. They take place in the context of urban redevelopment projects in which whole neighborhoods are razed and the vacant lots sold to real estate developers.[52] Besides residents, retailers constitute another group affected by evictions. Despite having a valid contract, they receive eviction notices requiring them clear their stalls within days. Others often operate without a license and, because they occupy space on sidewalks and thoroughfares, are frequently perceived as a nuisance by passers-by. Conflicts between the two groups are frequently documented in the (international) media, especially if conflicts lead to serious injury or death.[53] As with all such protests, the main issue of contention is the level of compensation, and, just like was case with protests land grabs, the cost of concessions is high and must be borne by the local government. It follows that repression rates should also be high.

5) Medical mistreatment, the reason for around six percent of all protests refers to grievances where doctors are accused of not having properly treated a patient. Such protests are frequently staged by relatives of individuals, often children, who have died in the care of physicians.[54] Tactics include bringing a coffin and arranging a wake near the entrance of the hospital where other patients and visitors can observe the spectacle. The relatives might demand compensation for their lost relatives and threaten to disrupt the operation of the clinic until the clinic meets their demands. Because hospitals are usually state owned, the amount of compensation handed out must be subtracted from the local budget. Once more, this issue type falls into the category where the cost of concession is high and where we should expect a high level of repression.

6) Around five percent of all protests are motivated by fraud-related issues such as investment scams. The recent years have seen an increase in fraud related to financial products, and defrauded investors take to the street in the hope that the government would step in to recover their investments.[55] Because investment scams are usually committed by private actors,[56] the cost of concessions is low for the government. Accordingly, the level of repression should also be low.

[50]Wedeman, Andrew, 'Not in my backyard: Middle class protests in contemporary China', in *The Middle Class in Emerging Societies: Consumers, Lifestyles and Markets*, ed. Leslie L. Marsh and Hongmei Li (Routledge, 2015), pp. 200–222.

[51]You-tien Hsing, 'The Great Urban Transformation: Politics of Land and Property in China' (Oxford: Oxford University Press, 2012).

[52]Eva Pils, 'Assessing evictions and expropriations in China: Efficiency, credibility and rights', *Land Use Policy* 58(2016), pp. 437–44; Cheuk Yuet Ho, 'Bargaining demolition in China: A practice of distrust', *Critique of Anthropology* 33(4) (2013), pp. 412–28.

[53]see for example this collection of articles in the South China Morning Post (SCMP, 'Chengguan', accessed January 27, 2020, http://www.scmp.com/topics/chengguan).

[54]Joseph D. Tucker et al., 'Patient physician mistrust and violence against physicians in Guangdong Province, China: A qualitative study', *BMJ Open* 5(10) (2015), accessed on January 27, 2020, https://www.ncbi.nlm.nih.gov/pubmed/26443652; Jiong Tu, 'Yinao: Protest and Violence in China's Medical Sector', *Berkeley Journal of Sociology* 11 (2014), accessed on January 27, 2020, http://berkeleyjournal.org/2014/12/yinao-protest-and-violence-in-chinas-medical-sector/; Benjamin L. Liebman, 'Malpractice mobs: Medical dispute resolution in China', *Columbia Law Review* 113(1) (2013), pp. 181–264.

[55]see for example 'How China's Peer-to-Peer Lending Crash Is Destroying Lives', *Bloomberg News*, October 2, 2018, accessed January 27, 2020, https://www.bloomberg.com/news/articles/2018-10-02/peer-to-peer-lending-crash-in-china-leads-to-suicide-and-protest; Pete Sweeney,'Breakingviews—Chinese peer lending protests set political trap', Reuters, August 7, 2018, accessed January 27, 2020, https://www.reuters.com/article/us-china-security-beijing-breakingviews/breakingviews-chinese-peer-lending-protests-set-political-trap-idUSKBN1KS0E7

[56]fraud by peer-to-peer lending platforms is especially common, see 'China Deploys Huge Police Force to Prevent Fraud Protest', *The Straits Times*, August 6, 2018, accessed on January 27, 2020, https://www.straitstimes.com/asia/east-asia/china-deploys-huge-police-force-to-prevent-fraud-protest

7) The last two protest issues, each of which accounting for roughly 2.5 percent of all protests, are less straightforward than the previous ones. As regards education, protesters might make demands that directly affect the local budget, an example being migrant workers demanding their children to be admitted into state schools.[57] However, protests might be directed against private schools who failed to hand out graduation certificates,[58] or stage by parents who are unhappy with the quality of food in the school cafeteria.[59] Hence, protests against public and private actors both fall into this category, and to ascertain which type of actor is targeted by a particular protest is difficult. Delineating the cost of concessions is not easier, because some demands are costlier than others. It is difficult to predict repression levels for this protest issue.

8) The situation is similar for environmental protest. The two main grievances in this category are the construction of electrical transformers in urban neighborhoods and garbage incinerators in the countryside.[60]

Urbanites often protest electrical transformers because they are concerned that these structures will lower the value of their property, while rural residents are worried about the negative impact of toxic fumes and contaminated water on their health. Theoretically, the plot of land designed for the electrical transformer could be used for residential purposes, so the cost of making concessions is low in financial terms. The matter is different for rural environmental protest, because land revenues and compensation payments from polluting industry might be the only chance for grass-roots officials to gain some revenue from otherwise inexpensive land.[61] In this case, the cost of concessions high. It follows that repression rate should be low for urban, but high for rural environmental protest.

Repression is reported by protesters to have taken place in 25 percent of all events, and it is remarkable that violence is more widespread than arrests. Events in which protesters report violence by state sanctioned agents constitute 19 percent of all cases. Arrests are reported in only 13 percent of all events. The data also provides insights into the identity of the perpetrators of violence. In her research, Lynette Ong demonstrates that local authorities might not use the police to quell unrest, but instead hire 'thugs'.[62] In their social media posts, some protesters indeed claim to have been roughed up by 'hooligans', 'ruffians', 'the mafia', 'thugs', 'punks' and so on. This is the case in 4 percent of all events. They are mentioned in 10 percent of all events in which violence also mentioned, which suggests that using hired thugs is a regular means of repression. Police presence (including armed police and traffic police) is mentioned in 15 percent of all events, urban law enforcement officials *(chengguan)* in 2.5 percent.

Correlates of Repression

This section uses the Wickedonna data to test the three theories introduced in chapter 2. To recall, one explanation regards repression as a function of protest intensity—the more intense a protest, the more likely repression becomes. A second explanation conceptualises repression as a threshold that imposes a price on collective action. The third explanation predicts that repression is more likely where the cost of concessions is high. This explanation is slightly modified by pointing out that each

[57]Yuanyuan Chen and Shuaizhang Feng, 'Access to public schools and the education of migrant children in China', *China Economic Review* 26 (2013), pp. 75–88.

[58]see for example Editor, 'Chinese police cracks down on student protests against education fraud', Followcn.com, March 14, 2019, accessed on January 27, 2020, https://www.followcn.com/chinese-police-cracks-down-on-student-protests-against-education-fraud

[59]Lily Kuo, 'Chinese parents storm primary school in rotten food row', The Guardian, March 13, 2010, accessed on January 27, 2020, https://www.theguardian.com/world/2019/mar/13/chinese-parents-storm-primary-school-in-rotten-food-row

[60]Graeme Lang and Ying Xu, 'Anti-incinerator campaigns and the evolution of protest politics in China', *Environmental Politics* 22(5) (2013), pp. 832–48.

[61]For an excellent book-length study on protests against waste incinerators, see Maria Bondes, *Chinese Environmental Contention. Linking Up Against Waste Incineration* (Amsterdam: Amsterdam University Press, 2019); see also Anna Lora-Wainwright et al., 'Learning to live with pollution: The making of environmental subjects in a Chinese industrialized village', *China Journal* 68(2012), pp. 106–24.

[62]Lynette H. Ong, 'Thugs and Outsourcing of State Repression in China', *The China Journal* 80(1) (2018), pp. 94–110.

issue is associated with grievance-specific costs of concessions for local governments. Instead of calculating the cost of concession for each individual event, the issue can be used as a proxy. To test this proposition, the article employs a logistic regression with repression as the dependent and protest issues and protest intensity (protest size and protest activity) as independent variables.

The event descriptions on the Wickedonna website report four kinds of protest activities: demonstrations, riots, collective petitions and strikes. At nearly 21 percent, demonstrations are the most frequent form of collective action in China. Demonstrations usually involve banner-carrying citizens shouting slogans in front of the city government, inside a gated community or in a factory. Collective petitions (17 percent) are the second-most frequent form of collective action. Collective petitioning is a form of shared action in which a limited number of citizens present formal demands to higher-level authorities. These demands tend to be more specific than those in demonstrations. In most cases, collective petitioners turn to higher-level authorities after lower-level authorities denied their request for (higher) monetary compensation. Nearly 14 percent of all protests turn into riots—participants overturn cars or destroy public property, but seldom attack officials or law enforcement officers. Strikes only occur in around one percent of all recorded events. To test for the impact of protest size, the natural log of the crowd estimates derived from the protest images is included in the model.

Apart from these variables, the model uses several controls. First, repression might be caused by structural features such as geography, leadership, or type of industry. Regarding labor protests, for example, Manfred Elfström and Sarosh Kuruvilla find that labor protests occur especially often in the Pearl River Delta, but also suggest that a distinct regional pattern is less pronounced than previously suggested.[63] To control for locality-specific fixed effects, dummy variables for each city are included in the model. Second, as repression might be a function of time, the model also tests for year fixed effects. Third, repression might not depend on protest issues, but whether a protest takes place in urban or rural areas. Arguably, repression is less costly in the countryside, because the danger of a protest spreading to other areas is lower. A 'countryside' dummy variable was constructed by

Table 2. Protest types and repression

	Model 1	Model 2	Model 3	Model 4	Model 5
Issues					
(constant)	0.298**	0.056**	0.056**	0.069**	0.113**
Labor	0.527**	0.684**	0.653**	0.642**	0.622**
Fraud	1.112*	1.071	1.067	1.083	1.08
Education	1.334**	0.994	0.936	0.919	0.865*
Medical	2.314**	3.001**	2.678**	2.584**	2.43**
Eviction	2.357**	2.622**	1.973**	1.883**	1.818**
Pollution	3.183**	2.521**	1.468**	1.444**	1.353**
rural_land	3.723**	4.178**	2.273**	2.181**	2.127**
Intensity					
Riot		1.666**	1.569**	1.505**	1.493**
demonstrate		1.301**	1.273**	1.217**	1.23**
Strike		0.762*	0.676**	0.652**	0.619**
Petition		1.717**	1.641**	1.646**	1.664**
log_size		1.353**	1.315**	1.326**	1.329**
Controls					
Village			2.332**	2.331**	2.389**
year FE				yes	Yes
city FE					Yes
Obs	71027	71027	71027	71027	71027
AIC	73462.91	68542.24	67245.14	67138.56	66528.74
BIC	73536.27	68661.46	67373.54	67294.46	69582.62
Pseudo R2	0.081	0.143	0.159	0.16	0.176

Note: ** $p \leq 0.001$, * $p \leq 0.05$. Reference category: labor protests. Cell entries display logit regression odds ratios.

[63]Manfred Elfstrom and Sarosh Kuruvilla, 'The changing nature of labor unrest in China', *ILR Review* 67(2) (2014), pp. 453–80.

searching for keywords like and 'village' and 'countryside'. 28 percent of all cases fall into this category.

Table 2 presents the results of the logistic regression. As each event belongs to exactly one issue, it is necessary to choose a reference category that all other categories relate to. Real estate protests seem to be an appropriate choice because its repression rate, 25 percent, is exactly the mean repression rate of the whole sample. Accordingly, the regression coefficients provide information not only on the relationship between an issue and repression relative to the reference category, but also to the average protest in the sample. For easier interpretation, the coefficients are reported as odds ratios (see Appendix C for the raw coefficients). An odds ratio below 1 indicates that someone who has encountered repression is less likely to have protested that particular issue than against real estate related grievances. This is the case for labor. Conversely, an odds ratio above 1 signal that an experience of repression is more likely to be associated with an issue, as is the case with protests fraud, issues related to education, medical mistakes or evictions. In the base model, the odds ratios for pollution and land requisition are above 3, meaning that protesters who have faced repression are more than three times more likely to have participated in a protest against pollution or landgrab than those who have not faced repression. Hence, a comparison of the coefficients reveals which issue is inherently riskier to protest against. As can be seen, the results largely confirm the hypothesis that repression is more likely to occur as the government's financial cost of concessions rises. It is surprising that the repression rate of homeowner protests is very near the average—the author assumed that homeowner protests would be unlikely to face repression. Possibly, this protest issue also needs to be further differentiated. Given the close relationship between local budgets, land sales and real estate development, it might be useful to control for state capture. Where there is little competition between developers, the latter might have a larger influence on politics, which would enable them to persuade local leaders to use repression against protesting homeowners.

Model 2 includes the protest intensity variables, i.e. size and protester activity. There are three observations worth noting. First, protest intensity does indeed seem to have an impact on repression. Repression is associated with a higher likelihood of riots, collective petitioning, and larger protests. Second, repression is independently associated with high cost of concessions and high protest intensity. Including intensity measures does not change effect size and statistical significance of most of the issues, and the predictive value of the model increases nearly two-fold. There are marked reductions in effect sizes only in the education and pollution issues. The association of education and pollution related protests with more intense form of collective action might be responsible for the decrease of their coefficients.

Similar findings hold for Model 3, which includes a control for protests taking place in the countryside. The effect of this variable is large, and its inclusion reduces the effect of the landgrab and pollution variables. The former remains significant and sizeable, but the latter remains barely above one. This indicates that environmental protests are primarily a rural, and not an urban phenomenon. It also suggests that many of the environmental protest recorded in the Wickedonna data set are the type of rural environmental protest that were associated with a high cost of concessions.

Including year fixed effects (Model 4) and city fixed effects (Model 5) has almost no impact on the results, indicating that protest intensity and cost of concessions affects repression independent of contextual factors such as geography or time. To account for the possibility that observations are correlated within cities, the models are recalculated with a generalized estimating equation (GEE) logistic regression and a random effects logistic regression. The results (not shown) are nearly identical.

The Likelihood of Repression

But what exactly do these coefficients tell us about the likelihood of repression? By themselves, they provide no information regarding how likely it is for protesters to encounter repression in each case.

If repression were a rare phenomenon, the likelihood to encounter repression might be low for participants in either type of event. To learn more about the actual likelihood of encountering repression, the logistic regression model that was fit to the Wickedonna data is used to make predictions on new data. By manipulating parameters such as protest issue, protester behaviour, location and crowd size, the predictions provide insight into how certain parameters or combinations of parameters increase or decrease the likelihood of repression. For a baseline prediction, all variables except for one issue variable, for example labor, are set to zero and the likelihood of repression is predicted if the number of participants is one, then two, and so on. This is done for a range between one and 10,000 participants. The same procedure is followed for all other issues. The baselines for each issue are visualized in the top left panel of Figure 1. The X axis indicates the number of participants, the y axis the predicted likelihood of repression. The results illustrate once more that repression is highly dependent on the underlying issue, and that protest intensity and the location of a protest are also strong predictors.

Figure 1 represents the values listed in Table 2 in a more intuitive manner. It clearly shows that some issues are more likely to encounter repression than others, and that added participants increases the likelihood of repression. Given the same number of participants, protests against medical mistreatment are most likely, labor protests least likely to encounter repression. Land grabs and evictions are also more, and protests against investment scams, housing issues and education less prone to be repressed. A single labor protester in an urban setting has an almost zero predicted likelihood of repression, but it is almost 20 percent for a protester who peacefully protests medical mistreatment in the same environment. As the number of participants increases, so does the risk of repression. As can be seen from the graph, if the same labor protest had 500 participants, the risk of repression would increase to 20 percent, equivalent to that of a single medical mistakes' protester. In contrast, an increase from one to 500 participants would triple the repression risk of a peaceful and urban medical mistakes protest to 50 percent. As the number of participants rises, the likelihood of repression increases logarithmically for all issues.

However, repression risk does not increase equally steeply for all issues. The steeper an increase, the more likely it is that the authorities will also move against very small protests. As expected, the curves are steeper for issues where the government's cost of concessions are high, and less steep than for issues where this is not the case. According to the predictions, a labor protest of 20 persons is only marginally more likely to meet with state-sanctioned violence or arrest than a sole migrant worker who displays his handmade cardboard protest sign before the gates of the shoe factory that owes him last months' wages. The situation is very different for protests medical mistakes. Here, one protester's predicted risk of repression is around 20 percent, that of 20 protesters 30 percent. This indicates that the government is relatively tolerant of labor protests and thus slower to intervene, but eager to nip protests in the bud where the cost of concessions is high. Here, even very small protests have a high risk of repression.

The other graphs in Figure 1 visualize the risk of repression for each issue if a protest is set in the countryside, violent, or both. The lowest line in each of the graphs represents the issue's baseline value, the one above repression risk when protesters riot, and the next one when a protest is set in the countryside. The highest line visualizes an issue's risk of repression when both conditions are satisfied: collective action takes the form of a riot in the countryside. Repression risk of a 500-person labor protest in an urban setting increases from 20 percent to 30 percent if participants apply violent means, and from 20 percent to 40 percent if it is staged in the countryside. The risk increases to 50 percent for a riotous labor protest with 500 participants in the countryside, nearly three times the likelihood of the baseline protest. For a medical mistake protest, the same variable constellation increases the risk of repression from 60 percent to 80 percent. These findings highlight the high level of contingency of repression in China. It confirms that certain issues have a higher risk of repression than others, especially if an issue is costly to resolve, and if the government would have to shoulder these costs.

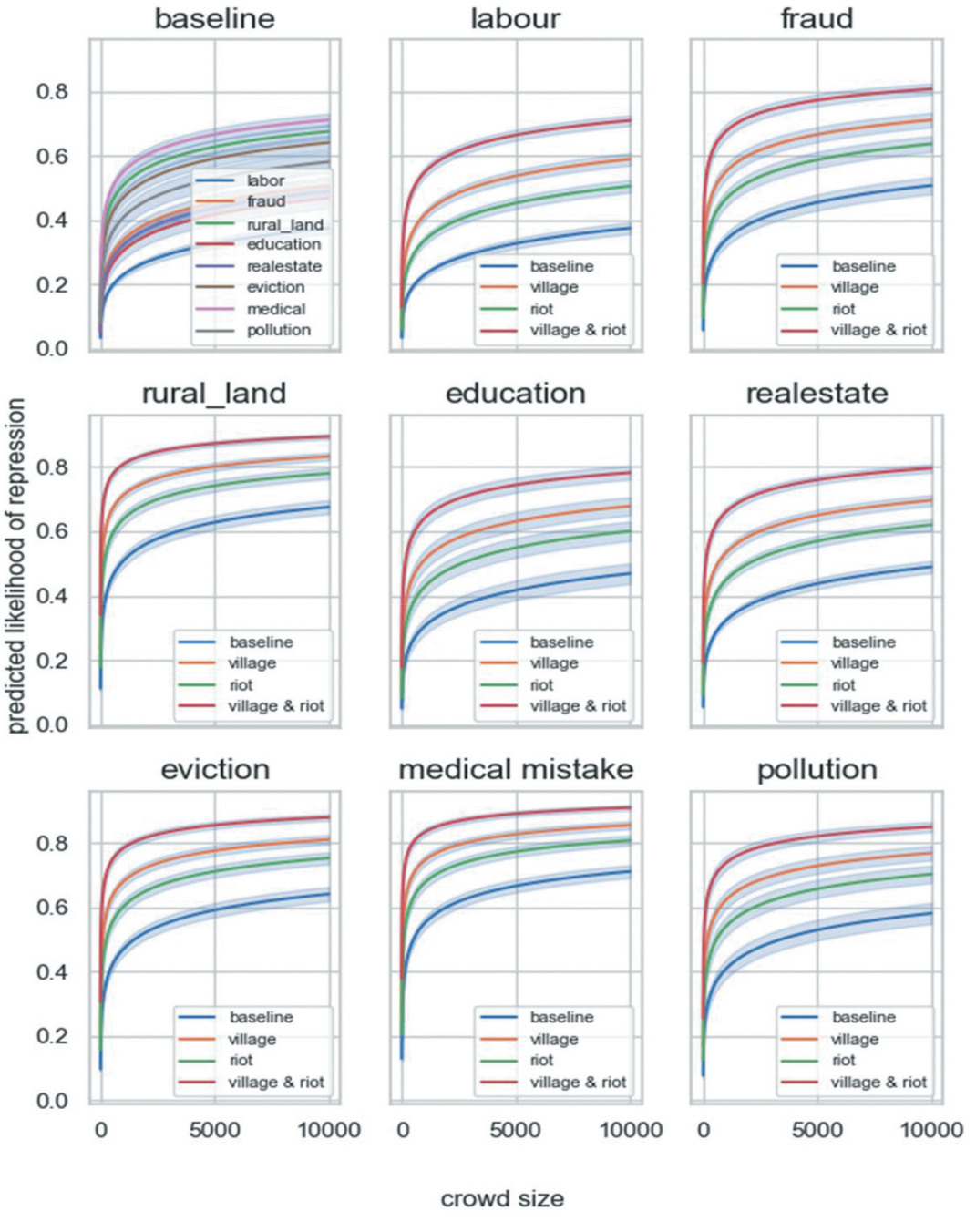

Figure 1. Predicted likelihood of repression: the effect of issue, location, activity and size. Note: Coefficients were obtained by using the logit model to predict the likelihood of repression for different combinations of issue, crowd size, location and protest activity.

Another interesting observation pertains to the distance between the baseline and the 'riot line' in each graph. The lower the distance, the less of a difference it makes for repression if a protest was peaceful or violent. Again, the distance is smallest for labor protests and highest for protests against medical mistakes. This might indicate that if workers take to the streets, officials opt for restraint even

if protesters apply violent means. The matter is very different for costly issues such as medical mistakes, land grabs and evictions, where a riot carries a 50 percent increased likelihood of repression. A possible explanation is that the authorities refrain from cracking down on a peaceful protest because they risk punishment by their superiors. When protesters use violent tactics, however, local officials are given a legitimate reason to intervene forcefully. This comes in handy especially where officials are looking for a reason, as might be the case in protests where concessions would be costly. In other words, a riot might just give them the excuse they need to crack down on a protest they are reluctant to resolve by making concessions.

The large effect of the countryside variable is also remarkable. The same protest is about twice as likely to face repression when it is set in the countryside as opposed to an urban setting. This confirms the assumption that an urban setting inhibits local officials from using violence against protesters, because an escalation is more likely in densely populated areas. In the countryside, however, protests are less likely to escalate and easier to contain. In terms of their population, rural regions are more homogeneous and less densely populated than city areas. Hence officials calculate not only the cost of concessions, but also the cost of repression. The former is dictated by the protest issue, the latter by a protest's location and the availability of a good reason to crack down on a protest.

Conclusion

Although there is a rich and expanding field of protest research on China, the determinants of government repression remain understudied. Existing studies all come to different conclusions about what causes repression and how frequently it occurs. Taking advantage of a new data set of more than 70,000 protest events sourced from social media, this article tested the validity of previous explanations of protest repression in China. It reveals that protest repression is not random, but predictable. If protests aim at financial concessions from local governments, repression is far more likely than if protests are directed against non-government entities.

It should be noted that the results presented here likely under-represent repression. Social media-based event data is biased against events taking place in localities where Internet penetration is low and residents do not have access to smartphones.[64] The fact that Internet penetration moderates protest frequency further complicates matters. Weidmann and Rød show that 'dictators use Internet-based communication to their advantage'[65]—online propaganda and surveillance both have a demobilising effect.[66] This could mean that protests are more likely to occur in remote areas with low Internet penetration, which are precisely the localities likely underreported in our sample. Yet it is in such localities that protests land expropriation and pollution often take place. This means that the real-world frequency of such protests is probably higher than reported in this study, and that repression occurs in many, if not most, of these events.

The article has raised some further issues future work must address. Among the challenges encountered are the disaggregation of protests related to education, real estate grievances and pollution, because their labels betray the complexity of the underlying issues. Case studies of such protests would be especially valuable.

Better insights also are needed on how the relationship between grievance and activity (riot, strike, protest) affects repression. Finally, while the analysis has shown that some protest issues are more likely to be repressed than others, there is still much room for improvement. For example, repression against labor protests is comparatively unlikely, but what are the circumstances that

[64]Nils B. Weidmann, 'A closer look at reporting bias in conflict event data', *American Journal of Political Science* 60(1) (2016), pp. 206–18; Christian Göbel and H Christoph Steinhardt, 'Better coverage, less bias: Using social media to measure protest in authoritarian regimes', *ResearchGate* (2019), accessed May 27, 2020, doi: 10.13140/RG.2.2.32856.75523/1.

[65]Nils B. Weidmann and Espen Geelmuyden Rød, *The Internet and Political Protest in Autocracies* (Oxford University Press, 2019), p. 154.

[66]Ibid, p. 144.

prompt officials to move against workers when they do? Similarly, when do local elites tolerate or make concessions to those who protest land grabs? What impact does protest leadership or the level of organisation of a protest have on the likelihood of repression? When do governments make concessions? When do they tolerate protests without making concessions? Here, Lorentzen's assertion that officials use random repression as a deterrent to prevent citizens with non-serious grievances from taking to the street offers a potential explanation.

These questions illustrate the limits of protest event data from social media. They are invaluable sources for learning about issues, frequencies and spatial and temporal distribution of social unrest, but provide little information about the internal dynamic of popular contention and about government responses apart from repression.

Acknowledgments

The research for this article was funded by the European Research Council (Grant No. 678266). The workshop 'New Evidence for the Study of Contention in China' was supported by the Chiang Ching-kuo Foundation for International Scholarly Exchange (Project Number: CS002-U-18). The author would like to thank Nils Weidmann, Kevin O'Brien, Meng Tianguang, Christine Ning, and Christoph Steinhardt for their comments on a previous version of the manuscript. Finally, this article would not exist if Lu Yuyu and Li Tingyu had not established the Wickedonna platform. For this, the author is deeply grateful.

Disclosure Statement

No potential conflict of interest was reported by the author.

Notes on contributor

Christian Göbel is professor of Modern China Studies at the University of Vienna, Department of East Asian Studies. A political scientist and sinologist by training, his research is concerned with the adaptability of the Chinese Party-State to social, economic and political challenges. He is interested in the effects of digital technology on local governance in China. He is the holder of a Starting Grant by the European Research Council to research the interrelationship between popular protest, online participation and public policy in China. His publications include books and articles on popular protest, e-government, digitalisation and urban and rural politics in China. Previous to his appointment in Vienna, he held positions in Duisburg, Lund and Heidelberg.

Appendix A: issue recoding

Lu Yuyu and Li Tingyu assigned keywords to more than two thirds of all events in their dataset. This appendix documents which of the original keywords were used to form the categories used in this analysis.

工人, 欠薪, 失业赔偿, 低薪, 克扣工资, 欠缴社保, 环卫工人, 关厂赔偿,民工, 讨薪: labor
业主, 物业, 租金: realestate
商业欺诈, 投资人, 融资诈骗, 储户, 规费: fraud
强征, 土地, 征地: land
强拆, 拆迁户, 搬迁赔偿, 拆迁: eviction
死者家属, 医疗事故, 医护, 病患, 医保: medical mistreatment
家长, 教师, 学生, 家属: education
出租车, 黑车, 车主, 公交车, 客运车, 三轮车, 运价, 运营权, 货运司机, 专车: transport,
贪腐, 官员贪: corruption
城管暴力, 警察暴力: police brutality

Appendix B: keyword dictionaries

This appendix presents the dictionaries used to measure repression (arrest or violence) and protest activities (demonstration, collective petitioning, riot, strike). The keywords were identified by means of word embeddings from a Word2Vec vector space model trained on the Wickedonna corpus.

arrest: 抓走, 抓捕, 逮捕, 捉拿, 抓到, 抓进去, 抓, 捉, 押, 抓去, 抓起, 逮, 逮到, 逮住, 拘留, 擒获, 抓有, 抓并, 拘禁,羁押, 刑拘, 抓名,带进, 扣压, 刑事拘留, 软禁, 捉去,带至, 扎, 抬走, 抓住, 捕捉, 捕获, 捕, 拘捕, 阻止, 拘, 非法拘禁, 关押, 行政拘留, 关进, 抓进, 抓入, 带入, 扣押, 抓获, 扣留, 拉走, 带走, 被关在, 抓往

demonstration: 示威, 散步, 静坐, 游行, 游行示威, 静坐示威, 示威游行, 行动, 活动, 集合, 游街, 集会, 举牌, 集聚, 聚众, 运动, 闹, 闹事, 闹起来

collective petitioning: 上访, 信访, 请愿, 诉求, 申诉, 投诉, 述求, 告状, 上诉, 上告, 表达, 状告, 申冤, 伸冤, 举报, 控诉, 起诉, 上访, 信访, 下跪, 跪下, 跪在, 喊冤, 苦苦哀求, 哀求, 哭诉, 诉苦, 喊冤

riot: 造反,抗议, 抗争, 反抗, 围攻, 斗争, 战斗, 讨伐, 抵抗, 对抗, 起义, 暴动, 闹革命, 暴乱, 发火, 走极端, 爆动, 翻天, 开战, 大闹天宫, 爆乱, 轰动, 打架, 群起, 奋斗到底, 不怕牺牲, 作死

strike: 摆工, 罢课, 罢市, 罢运

violence: 殴打, 打人, 伤人, 动手, 暴打, 开枪, 出手, 使用暴力, 开打, 打伤, 打死, 爆打, 欧打, 打晕, 乱打, 拳打脚踢, 施暴, 大打出手, 打得, 毒打, 砍伤, 打骂, 脚踢, 痛打, 砍杀, 拳脚相加, 击打, 拳打, 大打出手, 暴力行为, 袭击, 打死, 攻击, 挨打, 揍, 打打, 弄伤, 打成, 踢, 打坏, 打断, 打了个, 群殴, 围殴, 围打, 砍死, 推倒, 踹, 打至, 弄死, 捶打, 杀, 杀人, 杀死, 打掉, 烧死, 被打

Appendix C: Protest types and repression (raw coefficients)

	Model 1	Model 2	Model 3	Model 4	Model 5
Issues					
(constant)	−1.21**	−2.879**	−2.875**	−2.668**	−2.18**
	(0.018)	(0.036)	(0.036)	(0.043)	(0.059)
Labor	−0.641**	−0.379**	−0.426**	−0.443**	−0.474**
	(0.024)	(0.026)	(0.026)	(0.026)	(0.027)
Fraud	0.106*	0.068	0.065	0.08	0.077
	(0.045)	(0.047)	(0.047)	(0.047)	(0.049)
Education	0.288**	−0.006	−0.066	−0.084	−0.146*
	(0.055)	(0.061)	(0.062)	(0.062)	(0.063)
Medical	0.839**	1.099**	0.985**	0.949**	0.888**
	(0.035)	(0.037)	(0.038)	(0.038)	(0.039)
Eviction	0.857**	0.964**	0.68**	0.633**	0.598**
	(0.035)	(0.037)	(0.038)	(0.039)	(0.039)
Pollution	1.158**	0.925**	0.384**	0.367**	0.302**
	(0.053)	(0.057)	(0.059)	(0.06)	(0.062)
rural_land	1.314**	1.43**	0.821**	0.78**	0.755**
	(0.03)	(0.031)	(0.036)	(0.036)	(0.037)
Intensity					
Riot		0.51**	0.451**	0.409**	0.401**
		(0.027)	(0.027)	(0.027)	(0.028)
demonstrate		0.263**	0.241**	0.197**	0.207**
		(0.023)	(0.024)	(0.024)	(0.025)
Strike		−0.272*	−0.391**	−0.428**	−0.479**
		(0.087)	(0.089)	(0.089)	(0.092)
Petition		0.54**	0.495**	0.498**	0.509**
		(0.024)	(0.024)	(0.024)	(0.025)
log_size		0.303**	0.274**	0.282**	0.284**
		(0.007)	(0.007)	(0.007)	(0.007)
Controls					
Village			0.847**	0.846**	0.871**
			(0.023)	(0.023)	(0.024)
year FE				yes	yes
city FE					yes
Obs	71,027	71,027	71,027	71,027	71,027
AIC	73,462.91	68,542.24	67,245.14	67,138.56	66,528.74
BIC	73,536.27	68,661.46	67,373.54	67,294.46	69,582.62
Pseudo R2	0.081	0.143	0.159	0.16	0.176

Note: ** $p \leq 0.001$, * $p \leq 0.05$. Reference category: labor protests. Cell entries display raw coefficients. Standard Errors in paranthesis.