Article

The Reach of the State

Comparative Political Studies
2024, Vol. 57(8) 1243–1275

DOI: 10.1177/00104140231194057
journals.sagepub.com/home/cps
Sage

Charles Chang[1] **and Yuhua Wang**[2]

Abstract

We conceptualize the reach of the state by examining how the physical presence of the state helps the state project its power by signaling state interests and strength. We present a new measurement strategy to capture the territorial reach of the state using points-of-interest data provided by location-based service companies. Our measure exhibits several advantages: (1) it draws on firm-produced or crowd-sourced (rather than government-produced) data, (2) it includes highly precise, geo-referenced location information, which can be aggregated to any geographical or administrative level, (3) it traces temporal changes, and (4) it covers different types of state agencies. We illustrate its features using original databases that we compiled on state agencies in China and other countries. We demonstrate how researchers can use our measure by examining the locations and effects of coercive organizations and provide our data, code, and a tutorial to help researchers explore new avenues of inquiry.

Keywords

the reach of the state, state capacity, coercion, big data, China

[1]Duke Kunshan University, Kunshan, China
[2]Harvard University, Cambridge, MA, USA

Corresponding Author:
Yuhua Wang, Department of Government, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138, USA.
Email: yuhuawang@fas.harvard.edu

Data Availability Statement included at the end of the article

Heaven is high and the emperor is far away.

Chinese proverb

God is high and the czar is far away.

Russian proverb

Since ancient times, the Chinese and Russians have both demonstrated that they have figured out that the further away the state is, the more likely they are to get away with ignoring formal rules. This folk wisdom has inspired social scientists since at least Weber (1946 [1918], 78), who famously declared that territorial control is an essential component of state power. Mann (1984, p. 189) likewise argues that a state's "infrastructural power" allows it to implement its "political decisions throughout the realm." According to Shue (1988), the "reach of the state" determines how far it can penetrate society to exert control and regulate social relations.

We reinvigorate this concept by highlighting the importance of the physical presence of the state. We argue that state infrastructure, embodied in government buildings, bureaucratic offices, and police stations, influences citizen behavior by signaling state interest and strength. State infrastructure is therefore different from state capacity. The state decides to establish a physical presence in a locality not only because it has the *capacity* but because it also has an *interest* in doing so. A sizable political economy literature analyzes state territorial control as a strategic choice that maximizes fiscal extraction or political support at the lowest possible cost (e.g., Harbers & Steele, 2020; Herbst, 2014; Steinberg, 2018).

A large literature links a citizen's physical distance from the state to how they perceive its influence on their life. For instance, Stasavage (2010, p. 628) demonstrates that areas further from the state's political center have higher communication and transportation costs associated with political representation. Likewise, Brinkerhoff et al. (2018, p. 105) show that locations with fewer state facilities nearby experience lower levels of service utilization and satisfaction. Scott (2009, p. 167) similarly argues that people living in more remote locations that are less accessible to the state are culturally distinct from groups that are more accessible to the state. We build on this previous research to theorize about how the presence of state institutions helps the state project its authority.

Empirically, we introduce a new operationalization of the territorial reach of the state based on geo-referenced points-of-interest (POI) data related to state agencies provided by location-based service companies. Our approach helps researchers address four challenges associated with studying the state's territorial reach. First, our measure is based on firm-produced, street-level geo-survey and crowdsourcing data, which users update in real time. Firm-produced data

overcome some of the biases prevalent in government-released data, which are often manipulated or not publicly available (Hollyer et al., 2015; Wallace, 2016). Second, unlike other measures—such as census, taxation, or court cases, which are tied to certain administrative levels—our data are geo-referenced with highly precise geographical coordinates and can be aggregated to any geographical or administrative level or combined with other geo-referenced data (e.g., surveys). Our measure therefore avoids many measurement errors associated with aggregation (e.g., combining local tax accounts) and enables researchers to choose a theory-driven unit of analysis rather than be forced to focus on a particular administrative level due to data limitations (Soifer, 2019, p. 93). Third, while some popular measures of state reach, such as distance to the capital, do not vary over time, location-based service companies frequently update their databases, which enables researchers to track changes in state infrastructure. Lastly, our data cover various types of state agencies, including administrative, legal, coercive, and fiscal. Most prior empirical studies have focused on a particular state function, such as taxation. Our disaggregated measure will shed light on how the state allocates its infrastructure across multiple sectors (Fukuyama, 2013, p. 354).

We illustrate the features of our measure using an original database that we constructed by collecting over 1 million location points of Chinese government agencies from Amap.com—a location-based service company that has over 100 million daily users.[1] We demonstrate the reliability and validity of our measure using data from multiple years and different platforms and sources. We then compare our data with official statistics and show that missing information, caused by government officials' failure to release data on certain indicators, has systematically biased government statistics in China.

We then use two applications to demonstrate how our measure can help scholars answer interesting, sometimes unsettled, questions. First, we examine how a state locates its coercive agencies. We hypothesize that public spaces, such as parks and squares, provide "focal points" (Schelling, 1960, p. 57) for citizens to coordinate. In response, a government would establish more coercive agencies to monitor these public spaces to either preemptively or retrospectively deter social unrests. Analyzing fine-grained data at the township level in China in 2018, we show a strong positive correlation between the numbers of public spaces and police stations, even controlling for population and area.

Second, we use our data to shed light on contradictory empirical findings in the literature about whether state coercion increases or decreases dissent (or has no effect)—which Davenport (2007, p. 8) refers to as the "punishment puzzle." We argue that prior studies have reached different conclusions *partly* because many state repressive behaviors are invisible to the public. For example, harassment, surveillance, spying, bans, arrests, torture, and murder are focused on the targeted population and may have a deterrent effect on those who are directly repressed; their impact on the society at large depends

on the extent to which the general public is aware of these coercive behaviors. Our measure assesses the state's physical presence, which visibly signals its coercive capacity and intention to the public. We should therefore expect an increase in the presence of coercive agencies to lead to a decrease in contentious behavior. Exploiting a quasi-experiment in which the Chinese government rolled out a new type of coercive organization at the local level, and leveraging our data's exceptional temporal and geographic coverage, we use a difference-in-differences (DID) approach to show that counties in which these new coercive institutions were introduced in 2016 experienced significantly fewer mass protests in 2017 (compared with 2015).

Although our main illustrations use data from China, we demonstrate in the Online Appendix that researchers can obtain similar data for most other countries using our original global database that includes geo-referenced information on state agencies in 216 countries. We use data from OpenStreetMap, which provides open access data for the vast majority of countries. We validate this global measure by establishing a strong correlation among per capita state agencies, per capita government spending, and per capita public employees.

Our main contribution is to provide a fine-grained, data-driven measure of state reach.[2] Our approach complements an emerging empirical literature that employs a variety of novel methods to measure state activities (Brambor et al., 2020; Hanson & Sigman, 2021; Harbers, 2015; Hendrix, 2010; Lee, 2020; Lee & Zhang, 2017; Luna & Soifer, 2017; Soifer, 2015). Whereas states create institutions in an attempt to make society "legible" (Scott, 1998), our measure makes the state legible to citizens. In this sense, the concept underlying our measure is as much about the legibility of the state in the eyes of citizens as it is about the state's capacity to monitor and gather information about the citizenry. We therefore join a recent effort by Lee (2020, pp. 27–34) to use subnational variations in census quality to proxy for the reach of the state. We have made our data, Python code, and Structured Query Language (SQL) publicly available to facilitate future research.[3] We have also created an online tutorial to help researchers understand our code and data so they can apply our approach to their own contexts.[4]

Theoretical Motivations

Several prior studies have examined the spatial dimension of the state (Boone, 2003; Herbst, 2014; O'Donnell, 1993), in subfields including comparative politics (e.g., Slater, 2003; Ziblatt, 2006), international relations (e.g., Peic & Reiter, 2011), and American politics (e.g., Rogowski et al., 2022), and on a wide range of topics from Eastern European democratization (Ekiert, 1991) to American wartime mobilization (Tarrow, 2018). This research has established that a strong state presence promotes economic

development (Dincecco, 2017; Mattingly, 2017), prevents political violence (Kalyvas, 2006), and facilitates the delivery of public goods (Rothstein, 2011). In the study that inspired our measure, Soifer (2008, p. 242) argues that because the state has more power in some regions than others, scholars should examine the limited reach of radiating state institutions throughout the national territory.

Mann's (1984, 188–9) original framework treats state infrastructural power as a feature of the modern state. Feudal European states and imperial China had strong *despotic power*, defined as "the range of actions which the elite is empowered to undertake without routine, institutionalized negotiations with civil society groups," but weak *infrastructural power*, defined as "the capacity of the state to actually penetrate civil society, and to implement logistically political decisions throughout the realm." The state's infrastructural power has increased enormously in modern times, thanks to new technologies, expanded bureaucracies, and increased taxation, which enable the state to amass information about citizens and establish a presence in remote areas (Brambor et al., 2020). Mann (1984, p. 189) uses an analogy from *Alice in Wonderland* to describe state infrastructural power, which captures the (in)ability of the Red Queen to hunt down Alice: "Once you were out of sight of the Red Queen, she had difficulty in getting at you."

While Mann (1984) provides inspirations for our measure with his emphasis on the territorial reach of state power, state infrastructure is conceptually different from state infrastructural power. State infrastructure, represented by government buildings and bureaucratic offices, provides the resources for the state to exercise its power over society by signaling state intentions and reducing the transportation costs between the state and its population. In this sense, the difference between infrastructure and infrastructural power is similar to the difference between "inputs" and "outputs" of state capacity. While traditional works on state capacity often infer the level of state capacity from policy outputs and policy outcomes (e.g., tax-to-GDP ratio), as Brambor et al. (2020, p. 177) point out, it is preferable to measure state capacity by focusing on the "inputs," namely, the resources the state must deploy to realize its capacity. This is because there is a causal relationship between the resources invested and the outcomes obtained by the state. This relationship should be estimated rather than assumed. Similarly, the causal relationship between infrastructure and infrastructural power is an empirical question. As we will illustrate using the case of China, the establishment of coercive agencies successfully decreased the number of protests. In this case, state infrastructure enhances both the despotic and infrastructural powers of the Chinese state by facilitating the state's control of social actors.

Uneven Reach of the State

Where the state locates its agencies is a strategic action. Hotelling's model uses the metaphor of two ice cream vendors to illustrate how they strategically

set up their stalls on the beach: the (Nash) equilibrium location of the vendors should be right in the middle of the boardwalk because one vendor's deviation from the midpoint would allow the other to capture more than half of the market (Hotelling, 1929, pp. 45–51). Similar to setting up ice cream shops, the state strategically locates its agencies to economize its territorial reach. North (1981, p. 250) characterizes the relationship between citizens and the state as an exchange of revenue for services. The state trades a group of services, protection and justice in particular, for revenue. As a revenue maximizer, the state economizes the exchange by collecting the maximum level of taxation at the lowest cost (Levi, 1988, p. 3).

Prominent studies of state formation agree that the state extends its reach to grip taxation and population. Carneiro (1970, p. 734) argues that "pristine" states were more likely to emerge in densely populated, circumscribed areas where agricultural land was surrounded by mountains, seas, or deserts. Tilly (1992, p. 63) shows that European states began directly ruling their territories "as rulers bargained directly with their subject populations for massive taxes, military service, and cooperation in state programs." By contrast, Herbst (2014, p. 11) notes, "Relatively low population densities in Africa have automatically meant that it always has been more expensive for states to exert control over a given number of people compared to Europe and other densely settled areas."

State infrastructure, therefore, reflects not only state capacity but also state *intention*. Steinberg (2018) argues that even strong states have regions of limited state presence. Because all states have a finite amount of resources, strong states might strategically retain regions of limited state presence or subcontract state functions to non-state actors if they can enjoy greater political support by allowing a non-state actor (such as a firm, local chief, or non-governmental organization) to collect revenue on their behalf. Harbers and Steele (2020, p. 3) further challenge the assumption that a strong state must distribute its presence evenly throughout its territory. They recognize the potential benefits of heterogeneity: an uneven state presence makes it possible to accommodate distinct linguistic or cultural communities within a country and to tailor to various preferences for public goods provision in different localities. Territorial heterogeneity may thus be preferable to coerced uniformity.

Another reason why state presence is not uniformly distributed is that the state can exploit scale economies to govern more territory with fewer resources. The state enjoys economies of scale because there are fixed costs associated with establishing a set of facilities, such as government buildings, arsenals, and communication infrastructures. Up to a point, the costs increase less than proportionally to the population. To the extent that public services are non-rival and non-excludable, scale economies are achieved by exploiting these decreasing marginal costs (Alesina & Wacziarg, 1998; Friedman, 1977;

Gehlbach, 2008). The central state therefore needs fewer local branches on a per capita basis to govern a populated city than to administer a remote and less populated one.

How the State Regulates Behavior

The state projects its authority partly through its infrastructure.[5] When a local resident finds her lost wallet at the police station, a new driver gets her driver's license at the Department of Motor Vehicles, or a young couple applies for a marriage certificate at the city hall, the state shapes its interactions with citizens through its physical presence. In areas with a heavier state presence, bureaucrats are more likely to be embedded in local networks and to have more local knowledge (Bhavnani & Lee, 2018; Pepinsky et al., 2017)—both of which make the state more capable of monitoring the population, extracting taxes, maintaining order, and providing public goods (Deng & O'Brien, 2013; Koss, 2018; Lee & Zhang, 2017; Mattingly, 2016; Scott, 1998; Tsai, 2007; Xu & Yao, 2015).

A key theoretical insight that we use to derive testable hypotheses, which we later take to the data, is the role of state infrastructure in maintaining political order. When the state decides where to locate its coercive agencies, for example, an important consideration is to prevent potential disruptions. Space is inherently political. When individuals engage in public demonstrations, they do not simply occupy any random location; rather, they tend to gather in public parks, squares, or on prominent thoroughfares. These public spaces possess a distinct political character, as they serve as "focal points" that allow people to coordinate their actions (Schelling, 1960, p. 57). As such, the state may choose to station its coercive agencies within close proximity to these public spaces in order to monitor or deter or manage any large-scale unrest.

Once the state establishes its presence, it will increase citizens' perceived costs of joining an anti-state movement and suppress their incentive of initiating a protest, insurgency, or revolution. This link features most prominently in the literature on civil war. Fearon and Laitin (2003) examine a wide range of factors that have influenced the onset of civil wars since World War II and find that the most important predictor is state weakness—the inability of the state to control its territories. They argue that "to survive, the rebels must be able to hide from government forces," and "a distance from the centers of state power …should favor insurgency and civil war" (80). In a similar vein, Kalyvas (2006, p. 13) shows that the likelihood of violence during civil war is a function of state control: it is least likely to occur where the state is in full control.

Yet the importance of state presence extends beyond the ability to coerce the population. For North and Thomas (1973, p. 95), the key determinant of

economic growth in modern Europe was the emergence of state institutions (e.g., courts), which replaced local manors in governing social relations and protecting property rights. Another important function of the state is to deliver public goods, which are under-provided by non-state actors due to collective action problems (Olson, 1982, p. 19). A strong state presence facilitates the monitoring of individual behavior and helps curb free-riding in collective action dilemmas (Lee & Zhang, 2017, p. 120).

The State as a Unitary Actor?

Up to this point, our theoretical discussions have assumed the state to be a singular actor, consistent with the prevailing tradition in state theories (Evans et al., 1985; Levi, 1988; Tilly, 1992). However, the creation and implementation of state agencies involve multiple actors, as the decision-making power is not solely concentrated in a single entity. In a separation-of-powers system, such as the United States, the president can establish, divide, or eliminate federal agencies via presidential directives, subject to limited legislative oversight. Meanwhile, Congress can use its lawmaking authority to establish federal agencies and outline their fundamental operations (Huber & Shipan, 2002, p. 63). Moreover, subnational governments in most countries have discretionary power in their jurisdictions to create or abolish government agencies, a power that is more robust in federal than in unitary systems (Rodden, 2006).

While our assumption of a unitary state actor facilitates concise theoretical discussions, we advise contextualizing the inquiry in empirical studies. Depending on the vertical and horizontal divisions within the state, research must consider multiple actors whose decisions collaboratively shape the scope of the state. Our measure has a distinct advantage in this respect. As we explain in the following section, our data furnish insights into state agencies at different administrative levels and across diverse sectors, enabling researchers to disaggregate the state.

Gauging the Reach of the State

Previous studies have measured the state's territorial reach using road density (Goodwin, 1999; Herbst, 2014), taxation (Besley & Persson, 2009), court case records (Walker, 1999), or the location of primary schools (Soifer, 2015). Many studies use distance to the capital (e.g., Besley & Reynal-Querol, 2014) to proxy for state reach, assuming that state presence dissipates as one moves from the core to the periphery,

These traditional metrics present four challenges for empirical scholars. First, measures such as court case records rely on government-released statistics, which suffer from missing data and measurement errors. Lee and Zhang (2017) show that weak states lack the capacity to produce accurate statistics,

while Tsai (2008) and Wallace (2016) demonstrate that individual bureaucrats in authoritarian regimes have a strong incentive to "juke the stats" to advance their personal careers or serve broader political goals. Many governments, especially authoritarian ones, prefer not to disclose social, economic, and political data because transparency often generates political costs, such as increased levels of protest and unwanted exposure to the public (Hollyer et al., 2015; Malesky et al., 2012). If measurement errors or missingness in government statistics is correlated with state strength, estimates that use these measures as explanatory variables will be biased (King et al., 2001).

The second challenge relates to what Pepinsky (2019) calls "the return of the single-country study": political scientists increasingly leverage sub-national variations to test their hypotheses. As Soifer (2019, p. 93) argues, however, sub-national research designs are sensitive to the unit of analysis, and the risks of choosing the wrong unit are very high. Although the choice should be theory driven, scholars of sub-national politics often face a trade-off between sample size and data quality: the lower the level of government, the larger the number of observations, but the lower the quality of data. For example, national-level tax revenue data are widely available,[6] but fine-grained data on local taxation, especially in the Global South, are difficult to find (e.g., Touchton et al., 2021).

Third, the state is a strategic actor that allocates and adjusts its infrastructure to meet new challenges and achieve new governance goals. For example, various Chinese dynasties built and rebuilt the Great Wall and garrisoned their armies along the northern border to fend off threats from northern nomadic tribes (Lattimore, 1937). In the early 17th century, facing imminent threats from the European continent, Venice started to "expand its bureaucratic organs" in the mainland and "took costly measures to fortify" it against encroachment from other European powers (Ferraro, 2003, pp. 167–8, 191). While the dynamics of state territorial reach are important for understanding state intention, measures such as the distance to the capital often fail to capture temporal variations.

Lastly, state infrastructure varies across sectors, even within the same geographical unit (Skocpol & Finegold, 1982). Fukuyama (2013, p. 354) argues that because the state varies substantially across functions, researchers ideally seek "measures for *all* major government agencies" but warns that "this kind of data does not exist for many countries." Measures that focus on one dimension of state activity, such as road construction, taxation, court cases, or schools, do not allow researchers to examine how the state allocates its infrastructure across different sectors.

Our Measure

At the end of 2018, we used Python to collect POI data from Amap.com—a location-based service provider in China, owned by Chinese internet

conglomerate Alibaba.[7] Amap is one of the largest location-based service companies in the world; it provides hundreds of billions of locational services every day. By serving billions of mobile devices each month, it offers location-based services to major internet companies in China for ride sharing, food delivery, and social media. In addition, it provides mapping data for multinational corporations such as Google, Apple Maps, and Microsoft.[8]

Amap constructs its location database by combining geological surveying, street-level sensing and image processing, crowdsourcing, and user contributions. To maintain its maps' geographical accuracy, the company sends surveyors on a monthly basis to revisit locations and hires a large number of workers to digitize urban features into POIs that denote their geometric centers. It also integrates users' feedback to provide daily updates on geographical information and POIs.[9] As a result, Amap's POIs constitute the most competitive mapping service in China in terms of both volume and geographical accuracy. According to Chang's (2020) comparison of different location-based services in Kunming, the Amap database includes almost 500,000 data points, far more than Google Map's 70,428 data points.[10] To verify the database's accuracy, we examined approximately 100 state agencies with which we were familiar through our field work and found that they were all mapped at the exact correct locations with the exact correct agency names. Chang (2020, p. 465) more systematically calculates the geographic errors of different location-based services in Kunming and shows that Amap's average geographic error (the distance between the actual location and the map-designated location) is only 20 m, while Google Map's average geographic error is 394 m (Appendix Table A1-1).

The Amap database is not without limitations. For example, military institutions—for example, People's Liberation Army and People's Armed Police—are missing from its database because their locations are classified. Nor are politically sensitive buildings, such as "re-education camps" in Xinjiang, included. Since the database does not contain information on each organization's staff, we recommend using measures on organizational resources, such as the number of employees and budget size, to complement our measure. Lastly, Amap provides only the present version of its database and does not allow access to its archived data. Thus researchers must collect the data at least once a year to maintain an annually updated database.

We use the code provided by Amap (Appendix Table A1-2) to identify all government organizations, which produces 1,090,500 data points. Table 1 breaks down the state agencies included in the database by category.

Government administration, which constitutes over three-quarters of all state data points, includes the executive branch and its various functional departments (labor, education, environmental protection, etc.).[11] Legal institutions, including the police, procuratorate (i.e., public prosecutor), court, and notary, together comprise over 10% of the data points. Stability maintenance is a dispute resolution

Table 1. Categories of State Agencies in the Amap 2018 Database.

Category	Count	%
Government administration	840,947	77.116
Police	96,224	8.824
Procuratorate	10,725	.983
Court	17,450	1.600
Notary	29,476	2.703
Stability maintenance	29,200	2.678
Industry and commerce	23,779	2.181
Taxation	42,699	3.916
Total	1,090,500	100

institution established in the early 1990s to arbitrate increasing numbers of civil disputes. The police, along with stability maintenance, are also tasked with the coercive function of controlling and repressing social unrest. Industry and Commerce is a licensing agency that regulates economic activities. Lastly, taxation agencies collect taxes. We can roughly categorize these agencies into four major state functions: administrative, legal, coercive, and fiscal. Using principal component analysis, we show that all state agencies fall on a single dimension that explains over 85% of the variance (results reported in Appendix Figure A1-2).

Figure 1 (panel (a)) uses the total number of county-level (and below) state agencies in each county to rank all counties according to their percentiles.[12] Figure 1 (panel (b)) uses the per capita number of county-level (and below) state agencies in each county to rank all counties according to their percentiles.

The contrast between panels (a) and (b) is stark: while there are more state agencies in the more densely populated eastern part of the country, there are more state agencies per capita in the sparsely inhabited western areas. This is consistent with the discussion in the previous section that state infrastructure exhibits scale economies due to the decreasing marginal costs of governing a larger population. If a fixed number of state agencies are necessary to produce a public good that can be enjoyed by all residents of a region, then the per capita cost of producing that good—measured as state agencies per capita—will be less in more populous regions.

Figure 2 illustrates this tendency using a scatter plot that denotes the number of state agencies per million population (log) and population size (log) across Chinese counties. There is a clear negative relationship, which follows a power law: when the population increases by 1%, the number of state agencies increases by approximately .7% to keep pace.[13] Counties that lie above the regression line can be said to have a heavy state presence; those below have a comparatively low state presence.

The practical implication for empirical work that employs our measure is that if researchers use the number of state agencies (log) as an explanatory

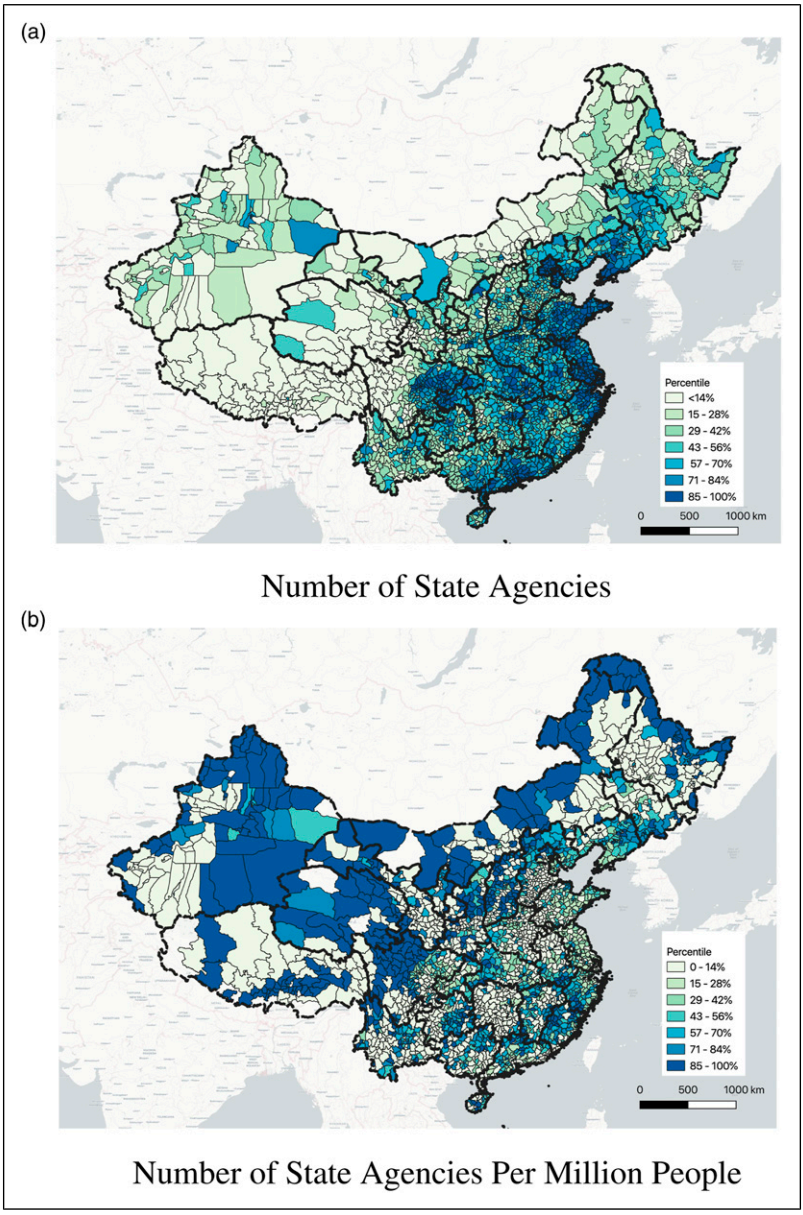

Figure 1. Density of state agencies across Chinese counties (2018).

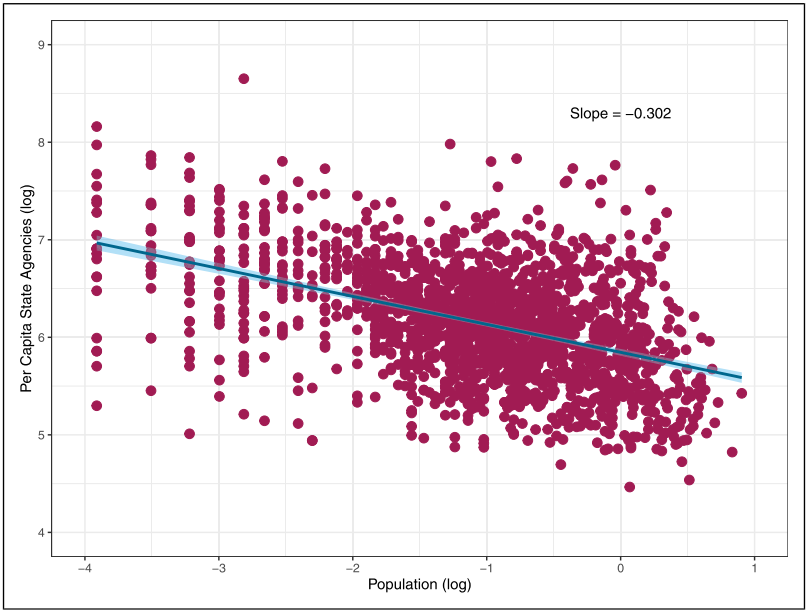

Figure 2. Economies of scale in state administration across Chinese counties.

variable, they should also include population (log) as a right-hand-side variable. The coefficient on state agencies measures the effect of state infrastructure, taking into account economies of scale in state administration.

Reliability and Validity

To assess our measure's reliability, we repeat our data construction procedure using data from multiple years and different sources. For other years, we use POI data from Amap that we collected in 2016 and 2017. Our alternative source is a similar database from Baidu Map that we collected in 2015.[14] To assess validity, we correlated the number of state agencies with the number of state employees, drawn from the 2010 census.

Figure 3 displays correlation plots between our measure—total number of state agencies from Amap 2018—and Amap 2017, Amap 2016, Baidu 2015, and total number of state employees in 2010. We aggregate all data to the county level—the lowest level at which state employee data are available. Our data are highly correlated (correlation coefficients >0.98) with Amap data from earlier years, which indicates temporal reliability. Our data also exhibit significant temporal variations: from 2017 to 2018, for example, 25,747 new state agencies were created. Our measure is also highly correlated with data from a different source and a different year: the correlation coefficient between

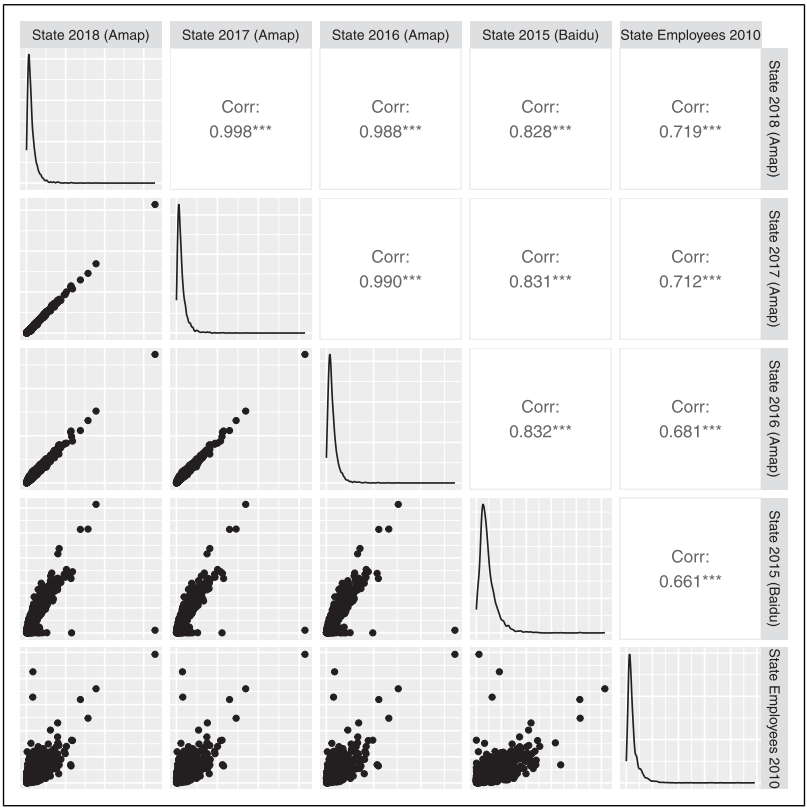

Figure 3. Correlations between Amap 2018 and data from other years and sources.
Note. The numbers of state agencies in 2016, 2017, and 2018 are all from Amap.com. The number of state agencies in 2015 is from Baidu Map. The number of state employees is from the 2010 Census.

Amap 2018 and Baidu 2015 is .83. Finally, we also discover a high correlation (correlation coefficient = .72) between our measure and the number of state employees.

A serious problem with government-released statistics is missing data. Appendix Figure A1-5 compares counties that have missing data on four important indicators—population, GDP, tax revenue, and fiscal spending—with those that released such data.[15] Counties that did not disclose these statistics had significantly more ($p < .01$) state agencies per capita than those that did.[16] As we elaborate in the Online Appendix, we believe that counties with more state agencies *choose* not to disclose these social and economic statistics because they want to avoid public scrutiny (Malesky et al., 2012) or avoid provoking protests (Hollyer et al., 2015).

Two Applications

We use two concrete examples to illustrate how researchers can leverage our data to address interesting, sometimes unresolved, questions about state power.

Public Spaces and Coercive Agencies

In the first exercise, we investigate how the state positions its coercive agencies. We assume that the state is a strategic actor that establishes these agencies to pre-empt or manage social unrest. A substantial body of literature has demonstrated that an individual's willingness to engage in social unrest depends on how many others will participate (Hollyer et al., 2015; Kuran, 1991; Lohmann, 1994). Consequently, social movements are coordination games. However, like the Battle of the Sexes, what allows citizens to form a shared belief in a unique equilibrium that enables coordinated behavior?

Previous works have focused on public information (Hollyer et al., 2015) or political events (Truex, 2019). Still, public spaces also play a vital role in coordinating people's expectations. As Arendt (1965, p. 31) wrote, "freedom itself needed …a place where people could come together – the agora, the market-place, or the polis, the political space proper." Public parks, squares, and main streets serve as focal points, enabling people to form their higher order beliefs—their beliefs about what other people believe regarding where to protest. Studying opposition protests in Putin-era Russia, for example, Armstrong et al. (2020, p. 17) show that "the vast majority of protests in regional cities occur on the main square." The Arab Spring, as Clarke and Kocak (2020, p. 1030) argue, started as "leaderless coordination of protester movements." Although it was not part of the original protest plan, marchers converged on Tahrir Square from around Cairo at the end of the day (Clarke & Kocak, 2020, p. 1030). Protesters converged on Tahrir Square because "Tahrir Square is in many ways an obvious place to protest" (Gunning & Baron, 2014, p. 245). And "[a]s the day progressed, square after square fell to the protesters" (Gunning & Baron, 2014, p. 2).

We propose a hypothesis that the number of public spaces within a certain territory is positively correlated with the number of coercive agencies, while holding population and area constant. We argue that the coexistence of public spaces and coercive agencies is part of multiple equilibria. In one equilibrium, a local government designates a piece of land as public space (such as a park or a square), possibly due to historical legacies or citizen demand. Anticipating an increased level of social protests in this public space, the local government also establishes more police stations to monitor the space. In this scenario, protests "off the equilibrium path" make the state conduct *preemptive* repression. In the other equilibrium, the emergence of public spaces leads to

actual protests, prompting the government to establish police stations to retrospectively control or prevent future protests.

However, determining causality between the establishment of public spaces and coercive agencies is challenging due to the possibility of government establishing police stations before or after opening public spaces. Additionally, our data limitations only allow us to conduct a cross-sectional analysis as we lack data on public spaces in other years for a panel analysis. Therefore, we focus on a correlational analysis while controlling for important confounders.

Measuring Public Space and Coercive Agency

To measure public space, we draw on our Amap 2018 database and include publicly accessible and free public spaces, such as local parks and public squares. We exclude sites that charge entrance fees, such as temples or United Nations Educational, Scientific, and Cultural Organization heritage sites. Entrance fees add costs to collective actions and make coordination less likely. Figure 4 (panel (a)) shows the number of public spaces at the township level (one level below county) across China.

To measure coercive agency, we use police stations, also drawn from the Amap 2018 database. The police is one of the most important social control mechanisms of the Chinese government (Wang, 2014). The police system is organized similarly to the government hierarchy, ranging from the national to the county levels, with police stations serving as the lowest units in the police hierarchy, established at the street (*jiedao*) or township level. The exact number of police stations in each street or township primarily depends on population size and area (Wang, 2014, p. 627). Figure 4 (panel (b)) shows the number of police stations at the township level across China.

Correlational Analysis

To test our hypothesis on how the state establishes police stations in relation to public spaces, we regress the number of police stations on the number of public spaces, both at the township level. We log transform all variables to consider scale economies of state institutions.[17] Table 2 reports the ordinary least squares (OLS) estimates.

In column (1), we include county-fixed effects to control for county-level covariates, such as leadership, geography, history, and fiscal conditions. We then add population size (log) in column (2) and area (log) in column (3) because population and area are primary criteria for establishing a police station. In all specifications, there is a statistically significant, positive correlation between the numbers of public spaces and police stations. Using the estimates in column (3), we can calculate that for every 1% increase in the

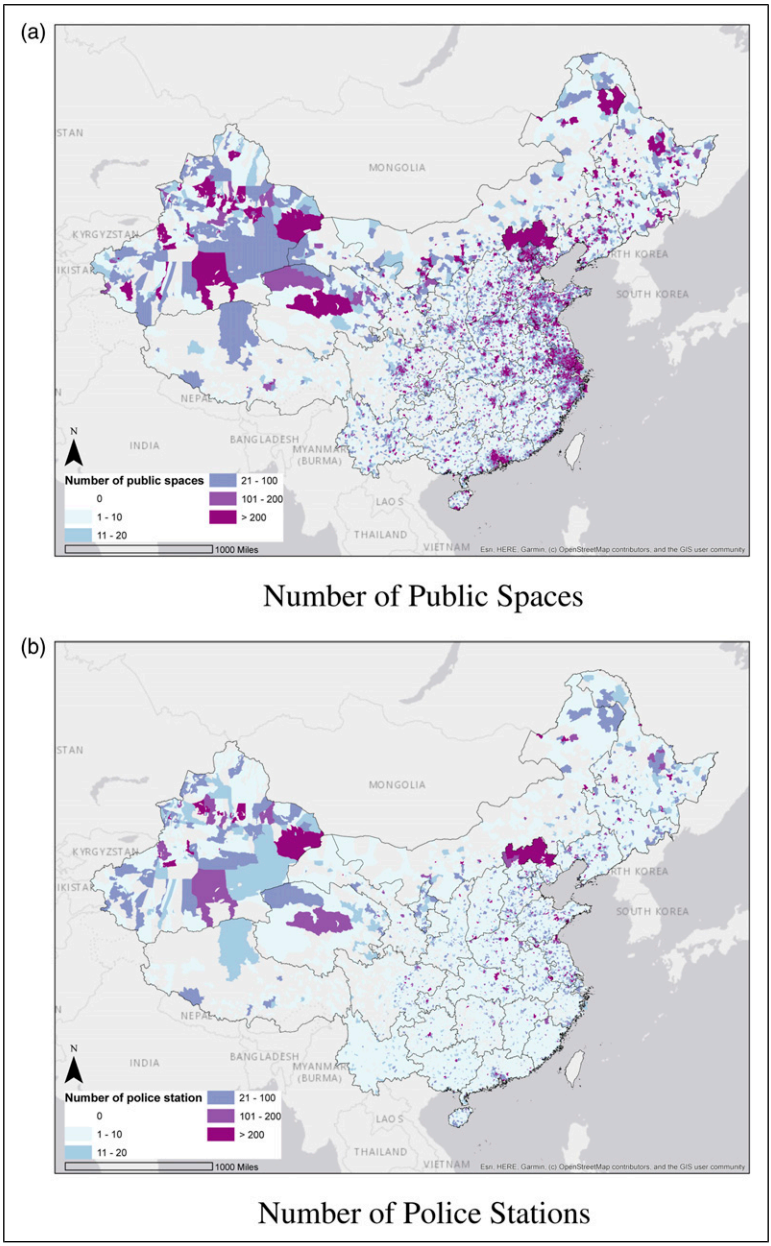

Figure 4. Public spaces and police stations across Chinese townships (2018).

Table 2. Public Spaces and Police Stations: OLS Estimates With Township-Level Data.

	N of police stations (log)		
Dependent variable:	(1)	(2)	(3)
N of public services (log)	.595***	.446***	.445***
	(.006)	(.006)	(.006)
Population (log)		.372***	.356***
		(.011)	(.011)
Area (log)			.056***
			(.009)
County F.E.	Yes	Yes	Yes
Observations	42,794	42,794	42,794
R^2	.894	.906	.907

Note. The unit of analysis is township. Standard errors are robust, clustered at the county level. *p*-Values based on two-tailed tests, *$p < .1$, **$p < .05$, and ***$p < .01$.

number of public spaces, the number of police stations increases by almost .5%. This effect is substantial: every new public square, for example, is associated with a 36% increase in the number of police stations.

In sum, we use our fine-grained data on public spaces and police stations to show that while public spaces provide "focal points" to coordinate mass behavior, the Chinese government has established more police stations to preempt or control social protests.

State Coercion and Protest

In the second exercise, we examine the relationship between coercion and dissent. Decades of prominent empirical work on state repression has yielded contradictory results. While many studies have suggested that repression undermines support for the government and heightens dissent by generating a backlash against the incumbents (Blaydes, 2018; Francisco, 1996; Lichbach, 1987; Rozenas et al., 2017), others have shown that repression reduces future opposition mobilization (Bellin, 2004; Hibbs, 1973; Levitsky & Way, 2012; Tilly, 1978; Wang, 2021). According to Carey (2006, p. 1), there is evidence available to support "almost every possible relationship" between repression and opposition.

A recent study by Pop-Eleches and Way (2021) seeks to account for such divergent findings by focusing on the visibility of state coercion. They argue that the impact of repression hinges on how much the public knows about the state's coercive behavior. Examining authoritarian repression of electoral protests in Moldova in 2009, Pop-Eleches and Way (2021) demonstrate that censorship moderates the relationship between repression and dissent: where

alternative media outlets are present, violence is more likely to increase opposition to the incumbents; where alternative sources of information are limited, repression may actually increase support for incumbents.

We build on Pop-Eleches and Way's (2021) insight and argue that *part* of the reason why previous studies have generated divided findings is that there are different types of state coercion. While tanks rolling over Tiananmen Square are highly visible to the public, they may not know about harassment, surveillance, spying, and bans, or the arrest, torture, and murder of political dissidents. Borrowing Scott's (1998) terminology, state coercion is more likely to reduce dissent when it is "legible" to citizens.

Motivated by our earlier theoretical discussions, we argue that the physical presence of coercive organizations depresses dissent through two mechanisms. First, the establishment of a new coercive organization signals that the state has both the capacity and intention to punish dissent in the neighborhood. These signals enter into citizens' *ex ante* calculations when deciding whether to join a protest: expecting the state to respond with repression, risk-averse citizens would not participate in the first place. In a second mechanism, the availability of coercive facilities will enable the state to react quickly if a protest happens. The state, like the Red Queen, needs to be geographically close to the protest to keep it under control. Even if state institutions are equipped with digital surveillance capabilities, dispatching police cars to the protest site takes time, which is positively correlated with the distance between the protest and the nearest police station. The existence of coercive facilities will influence how a protest unfolds *ex post*: a nearby police station is more likely to stop the protest from escalating than a far-away one.

We therefore hypothesize that after a locality establishes a coercive organization, protests are less likely to happen. To estimate the effect of coercive organizations on protests, we exploit a quasi-experiment in which the Chinese government rolled out local-level coercive agencies; we leverage the exceptional regional and temporal variations in our data on state agencies to assess the impact of this policy.

Stability Maintenance

Stability maintenance is a euphemism the Chinese Communist Party (CCP) uses for social control. A key stability maintenance agency is the Committee for Comprehensive Management of Public Security (CMPS), which was created in the early 1990s as a policy response to the Tiananmen democracy movement and the collapse of communist political systems in Eastern Europe (Wang & Minzner, 2015, p. 349). Unlike government agencies that specialize in coercion, such as the police or the military, the CMPS coordinates government responses to social unrest through a broad range of government and social organs. The committee's work goes beyond simply mobilizing law enforcement to arrest, prosecute, and jail offenders; it also involves making

individual workplaces or schools responsible for controlling their students or workers, and tailoring responses to the circumstances of each individual (Wang & Minzner, 2015, p. 346).

In 1991, the CCP established a central CMPS at the national level. Over the next two decades, the CCP ordered the creation of CMPS branches at the county level and higher (Wang & Minzner, 2015, p. 349). In 2013, the CCP required CMPS branches to be established at more local levels (below county).[18] A CCP document from 2015 ordered the creation of a CMPS branch in every county, township (one level below county), and village (two levels below county).[19] 2016 was a critical year for the rollout of local CMPS branches. According to a report by China's Xinhua News Agency, by the end of 2016, 97% counties had established a three-level CMPS system.[20]

Our data on state agencies in 2015 and 2016 allows us to capture the new stability maintenance agencies established in 2016. We find that about one-third of Chinese counties established at least one new stability maintenance agency in 2016 (Figure 5).

Protest

Similar to previous research on contentious behavior, our outcome variable is the number of protests. The main data source used to create collective action datasets for protest event analysis has been traditional media, particularly newspapers and newswire press releases (e.g., McAdam & Su, 2002), but newspapers are an imperfect source because their coverage is likely to over-represent larger and more sensational protests (McCarthy et al., 1996). In authoritarian countries, it is especially challenging to use traditional media as a target source to study protest events because the regimes impose strict controls on news reporting through state ownership of media outlets (Stockmann, 2013). As a consequence, many protest events that take place in China are not reported in traditional media.

We use a dataset recently made available by Zhang and Pan (2019), who use convolutional neural networks on image data and recurrent neural networks with long short-term memory on text data in a two-stage classifier to identify social media posts about offline protest events. They used this approach to analyze Chinese social media data and identified more than 100,000 protest events from January 1, 2010 to June 30, 2017. Zhang and Pan (2019, p. 34) evaluate the performance of their approach through cross-validation, out-of-sample validation, and comparisons with other protest datasets and conclude that their data have a broader coverage and are more accurate than most datasets based on media reports. In particular, they assess the effect of online censorship and find that it does not substantially limit their identification of events. Zhang and Pan's (2019) data are not without limitations. Compared to other protest datasets, Zhang and Pan (2019) identified

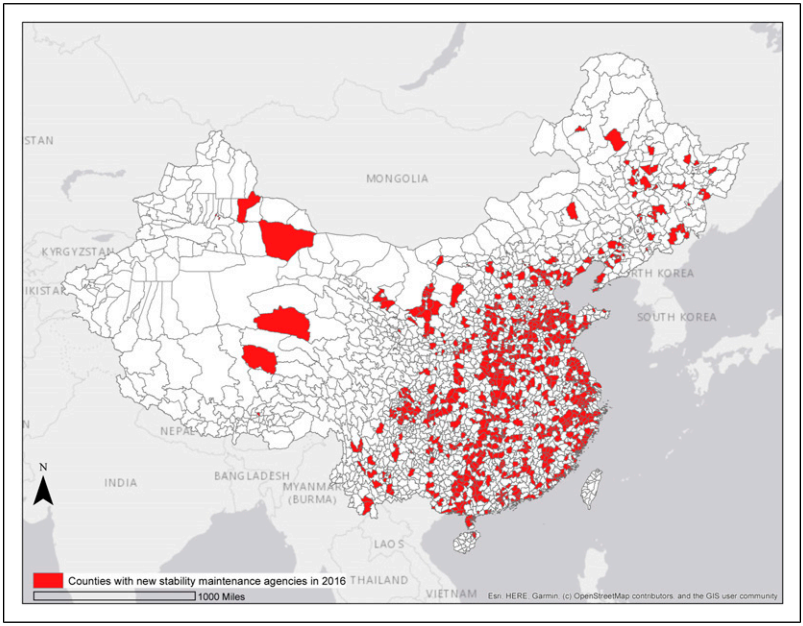

Figure 5. Counties with new stability maintenance agencies in 2016.

more rural, land-related protests, and relatively few protest events related to ethnic and religious conflict. But given the difficulties of accurately measuring protest events in authoritarian regimes, their dataset is the best available.

Because our treatment variable—new stability maintenance agencies—was measured in 2016, we are interested in how the number of protests changed from 2015 to 2017. Since Zhang and Pan's (2019) data include only the first half of 2017, we compare it with the first half of 2015. Figure 6 shows the number of protests across Chinese counties in the first halves of 2015 (panel (a)) and 2017 (panel (b)).

Difference-in-Differences Estimates

We employ a DID strategy to estimate how establishing stability maintenance agencies in 2016 changed the number of protests between 2015 and 2017. The *first difference* is the temporal difference: how the number of protests across Chinese counties changed from 2015 to 2017. The *second difference* is the regional difference: how the number of protests varied due to the establishment of new stability maintenance agencies. The DID design, therefore, can identify the differential effects of establishing a new coercive organization across counties.

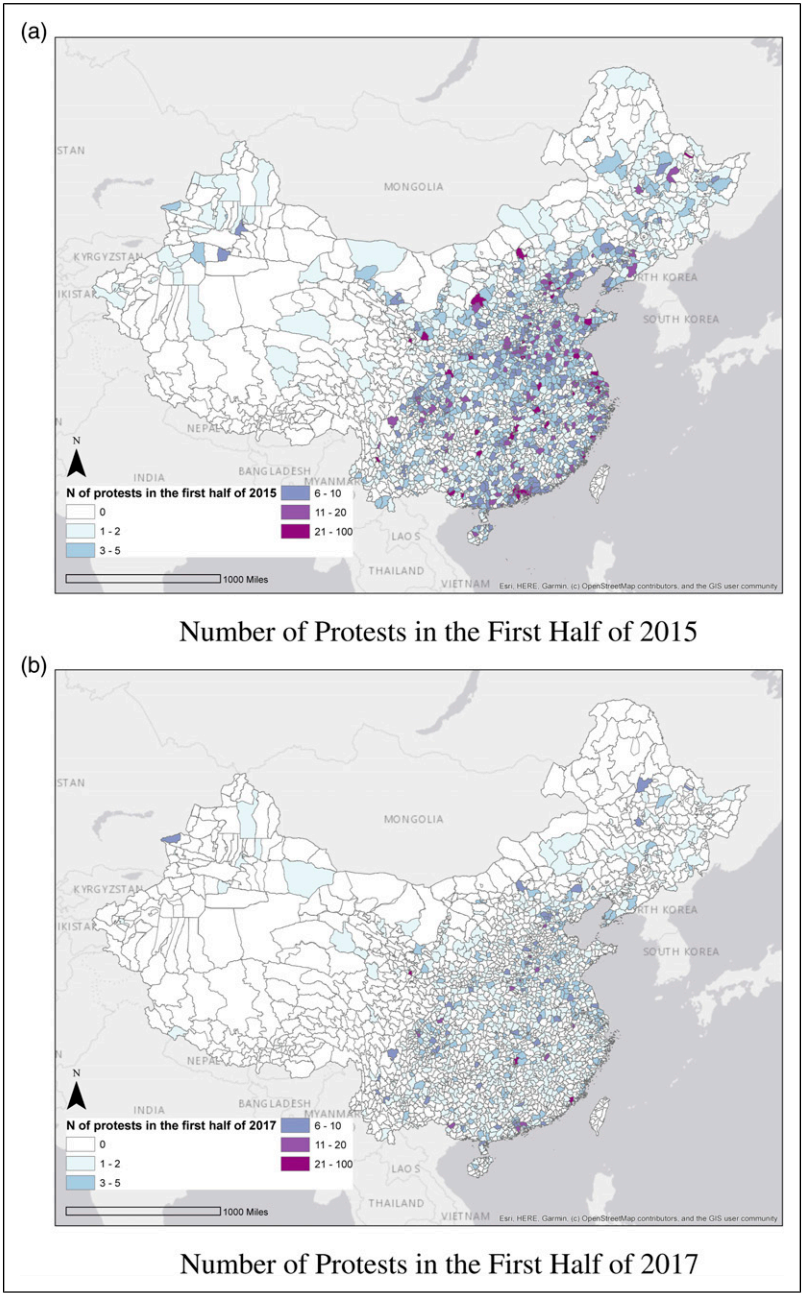

Figure 6. Number of protests across Chinese counties in 2015 and 2017.

The treatment group consists of counties that established at least one new stability maintenance agency in 2016; those that did not constitute the control group. Before 2016, the number of protests in the control group is denoted by Y_1^C; after 2016 it is Y_2^C. The number of protests in the treatment group is Y_1^T before 2016, while afterward it is Y_2^T. The DID estimator is defined as follows:

$$DID = \left(Y_2^T - Y_2^C\right) - \left(Y_1^T - Y_1^C\right) \quad (1)$$

The identification assumption is that, if the stability maintenance agencies had not been rolled out, the difference in the number of protests between the control and treatment groups would have been constant over time. While we cannot directly test this common trends assumption, if we have more than one pre-treatment period for which data are available, pre-existing differences in the trends of the outcome variable can be detected by applying the DID estimator to pre-treatment data (Abadie, 2005, p. 2).

To evaluate the credibility of the common trends assumption, we compile a panel dataset of Chinese counties across three years (2013, 2015, and 2017). We include a "lead" in our regression to test whether the changes in the number of protests from 2013 to 2015 (pre-treatment period) differed in the control versus treatment groups. If the DID assumption is plausible, the coefficient on the "lead" should be zero. Appendix Table A1-3 reports the results of this regression. We include county-fixed effects to control for county-level time-invariant factors such as geography, culture, and history. We also include year dummies to account for widespread shocks specific to each period, such as population growth, national policy changes, and macroeconomic conditions. All standard errors are robust, clustered at the county level to account for any within-county serial correlation in the error term. The coefficient on the "lead" (Year2015×Treatment) is small (−.301) and indistinguishable from zero, while the coefficient on the DID estimator (Year 2017×Treatment) is large (−1.384) and statistically significant at the .01 level.

Figure 7 displays the standard DID plot of the changes in the number of protests from 2013 to 2017 in the treatment and control groups. While the average number of protests in both groups followed a parallel trend from 2013 to 2015, they diverged after 2015. The dotted line traces the trajectory of the counterfactual scenario in which no new stability agencies were introduced in 2016. The treatment group diverged from the counterfactual to experience 1.384 fewer protests. This effect size is substantial: the average number of protests is 2.672, and establishing a stability maintenance agency would reduce the number of protests by half of its mean (and over a quarter of its standard deviation). Our results indicate that establishing coercive organizations effectively reduced protests.

We also conduct a placebo test to examine whether the reduction in protests was particularly due to the establishment of coercive organizations or an

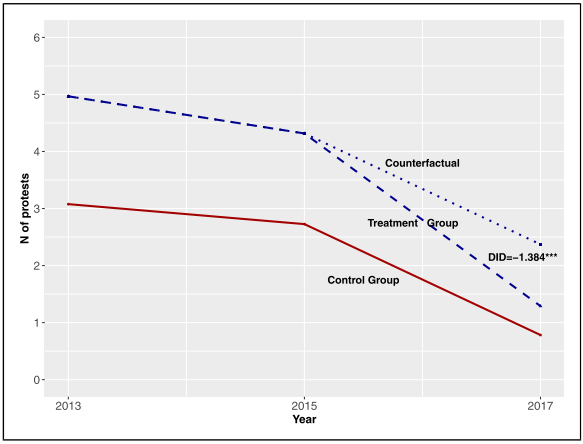

Figure 7. DID Plot.
Note. The treatment group is defined as Chinese counties that established at least one stability maintenance agency in 2016; the control group contains those that did not. The data include Chinese counties across three years (2013, 2015, and 2017). The *y*-axis indicates the average number of protests. Appendix Table A1-3 shows the estimates.

increase in state presence in general. Exploiting the rich information in our data, which includes all kinds of state agencies, we use the same data construction methods and model specification in Appendix Table A1-3 to examine how an increase in administrative, legal (e.g., court), licensing (e.g., industry and commerce), and taxation agencies influenced protest. As Appendix Table A1-4 shows, the establishment of these non-coercive organizations does not exert a significant effect on protest. While all of the coefficients are negative, they are all small and indistinguishable from zero. The results of this placebo test indicate that the reduction in protests from 2015 to 2017 was specifically due to the establishment of coercive agencies and not to an increased presence of the state more generally.

In sum, we leverage our data's regional and temporal variations and exploit a quasi-experiment in which the Chinese government installed local-level coercive organizations in 2016 in some counties to demonstrate that increasing the physical presence of a coercive state can effectively decrease the frequency of protests in an authoritarian regime.

Conclusion

After World War II, a large number of countries declared independence from their colonial powers and established their own states. Almost all of these newly independent states emerged in what is now known as the developing

world: sub-Saharan Africa, the Middle East, Asia, and Latin America. Many of these countries have struggled to extend the reach of the state beyond the capital. Crimes, poverty, and civil wars occur to a large extent either because the state cannot penetrate society or after it loses its control over territories. Building up the state has hence become the top priority of these countries and the international community. Fukuyama (2004, p. 18) argues that state building (rather than promoting democracy and development) is the key goal in international development.

The reach of the state has been a much invoked concept in the social sciences. We reinvigorate this concept here by highlighting the importance of the physical presence of the state. The existence of government buildings, bureaucratic offices, and police stations signals state capacity and intention, and facilitates state regulation of citizen behaviors. We illustrate this argument by showing that establishing coercive organizations on the ground decreases mass protests at the local level in China. Our empirical analysis suggests that a crucial component of state power comes from the state's visibility to citizens.

The study's most important contribution is the production and dissemination of our state infrastructure dataset. While our primary illustration uses the case of China, our measure can be used to conduct research on state infrastructure in a large number of countries in which official statistics have been opaque or difficult to obtain. In the Online Appendix (p. A13), we discuss how our measurement strategy can create a cross-national database using POIs from OpenStreetMap. Our measure therefore provides excellent geographic and temporal coverage. This feature permits a level of flexibility in specifying the unit of analysis that will be useful for scholars of comparative politics and beyond. Additionally, our measure based on OpenStreetMap can be calculated for historical periods (as early as 2005) and future years as well. To the best of our knowledge, our state infrastructure indicator has the broadest cross-national, sub-national, and historical coverage available.

We also stress that, unlike similar indicators, a key feature of our measure is that it does not depend on government statistics, which are often manipulated. Relatedly, compared with many existing measures that focus on the "outputs" of the state, such as taxation, our measure reveals the "inputs" of the state by gauging the resources the state has invested. This enables researchers to empirically estimate the relationship between state investment and capacity realization, which has often been assumed in prior works. Many recent efforts to measure state activities use an index, which is often estimated using a latent variable approach (e.g., Hanson & Sigman, 2021). While our measure is correlated well with existing measures (Appendix Figure A1-9) and has the same advantage of covering state activities in different dimensions, it has an extra reward of facilitating interpretations. Researchers can precisely interpret their results by estimating the effect of having one more police station or five more tax bureaus.

We believe our data can open up novel avenues for research on the state's territorial reach. To facilitate future research, we have made our data, Python code, and SQL publicly available.[21] We have also provided a short tutorial for researchers to access and use our data on Google BigQuery.[22] Our data include POIs from both China and other countries. Our code enables researchers to collect data from a variety of location-based services, such as Amap, Baidu, Google Maps, Tencent, and OpenStreetMap. A renewed focus on state infrastructure has the potential to yield novel theoretical and empirical insights into a range of outcomes of interest to political scientists; we have only scratched the surface here.

Acknowlegments

We greatly thank Pablo Beramendi, Melissa Lee, Imke Harbers, Daniel Koss, Melanie Manion, Dan Mattingly, Carl Müller-Crepon, Hillel Soifer, Jeremy Wallace, Xu Xu, and participants at APSA 2021 Chinese Politics Mini-Conference and the Humanities Research Center and Center for the Study of Contemporary China Faculty Research Workshop at Duke Kunshan University for comments. All errors remain our own.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Charles Chang https://orcid.org/0000-0002-6518-9014
Yuhua Wang https://orcid.org/0000-0003-4311-5425

Data Availability Statement

The data given this article are available at https://doi.org/10.7910/DVN/CSBRZX.

Supplemental Material

Supplemental material for this article is available online.

Notes

1. https://rb.gy/xzilbi (accessed August 5, 2020).
2. Replication materials and code can be found at Chang and Wang (2023).

3. All of our data and code can be downloaded here: https://github.com/placeasmedia/stateagencies.
4. https://github.com/placeasmedia/stateagencies/blob/main/sql_bigquery/Access_POI_Data_Using_BigQuery.ipynb.
5. While the internet allows the state to regulate citizen behavior without having a physical presence (e.g., through internet censorship, see King et al. (2013)), many state functions still require a physical location.
6. See, for example, https://data.worldbank.org/indicator/GC.TAX.TOTL.GD.ZS (accessed October 26, 2022).
7. Application Programming Interface (API) can be used to request access to Amap's database and submit up to 30,000 API requests per day for free with a registered account. More records are available for a fee. The Appendix describes our data collection process.
8. https://rb.gy/wibbce (accessed August 5, 2020).
9. While we focus here on geocoded government buildings, there are many other potential applications. For example, Amap's real-time congestion POIs from smartphone users can display protests or crowd gatherings in real time.
10. Google Maps provides fewer POIs in China than in other countries; the company has not kept up with POI digitization since its retreat from China in 2010 after a dispute over internet censorship (Chang, 2020, p. 466).
11. Amap also indicates the territorial level at which the government agencies are administered (national, provincial, prefectural, county, and township).
12. Our data allow us to examine state agencies at different administrative levels. Appendix Figure A1-3 illustrates that different levels of state agencies are weakly correlated, suggesting that researchers should treat state agencies at different levels as separate variables. The use of these variables in empirical analysis, consistent with Soifer's (2019, 93) call to be mindful of the unit of analysis, will produce vastly different empirical results. Appendix Figure A1-4 exploits the geo-referenced nature of the data and illustrates the different state agencies as points in the political center of the country—Dongcheng District in Beijing.
13. A regression of log state agencies per capita on log population with an estimated coefficient of −.3 is equivalent to a regression of log state agencies on log population with an estimated coefficient of .7.
14. Baidu Map is another web mapping service, provided by Baidu—a Chinese technology company specializing in internet-related services and products. Baidu Map has over 300 million users in China and provides map data for multiple countries and multinational companies, including Tesla. See https://rb.gy/nm2dty (accessed August 5, 2020).
15. We rely on statistical yearbooks published by the Chinese government at various levels to measure whether a county has released such data.
16. We use the 2010 census data, which is more complete, to calculate per capita state agencies.
17. For example, N of Police Stations (log) = log(N of Police Stations +1). We add 1 to deal with townships with zero police stations.

18. http://www.chinapeace.gov.cn/chinapeace/c53712/2016-06/06/content_11729466.shtml (accessed October 27, 2021).
19. http://politics.people.com.cn/n/2015/0414/c1001-26839083.html (accessed October 27, 2021).
20. http://www.xinhuanet.com//legal/2017-09/17/c_1121677387.htm (accessed October 27, 2021).
21. All of our data and code can be downloaded here: https://github.com/placeasmedia/stateagencies.
22. https://github.com/placeasmedia/stateagencies/blob/main/sql_bigquery/Access_POI_Data_Using_BigQuery.ipynb.

References

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, *72*(1), 1–19. https://doi.org/10.1111/0034-6527.00321

Alesina, A., & Wacziarg, R. (1998). Openness, country size and government. *Journal of Public Economics*, *69*(3), 305–321. https://doi.org/10.1016/s0047-2727(98)00010-3

Arendt, H. (1965). *On revolution*. Penguin.

Armstrong, D., Reuter, O. J., & Robertson, G. B. (2020). Getting the opposition together: Protest coordination in authoritarian regimes. *Post-soviet Affairs*, *36*(1), 1–19. https://doi.org/10.1080/1060586x.2019.1665941

Bellin, E. (2004). The robustness of authoritarianism in the middle east: Exceptionalism in comparative perspective. *Comparative Politics*, *36*(2), 139–157. https://doi.org/10.2307/4150140

Besley, T., & Persson, T. (2009). The origins of state capacity: Property rights, taxation, and politics. *The American Economic Review*, *99*(4), 1218–1244. https://doi.org/10.1257/aer.99.4.1218

Besley, T., & Reynal-Querol, M. (2014). The legacy of historical conflict: Evidence from africa. *American Political Science Review*, *108*(2), 319–336. https://doi.org/10.1017/s0003055414000161

Bhavnani, R. R., & Lee, A. (2018). Local embeddedness and bureaucratic performance: Evidence from India. *The Journal of Politics*, *80*(1), 71–87. https://doi.org/10.1086/694101

Blaydes, L. (2018). *State of repression: Iraq under saddam hussein*. Princeton University Press.

Boone, C. (2003). *Political topographies of the african state*. Cambridge University Press.

Brambor, T., Goenaga, A., Lindvall, J., & Teorell, J. (2020). The lay of the land: Information capacity and the modern state. *Comparative Political Studies*, *53*(2), 175–213. https://doi.org/10.1177/0010414019843432

Brinkerhoff, D. W., Wetterberg, A., & Wibbels, E. (2018). Distance, services, and citizen perceptions of the state in rural africa. *Governance*, *31*(1), 103–124. https://doi.org/10.1111/gove.12271

Carey, S. C. (2006). The dynamic relationship between protest and repression. *Political Research Quarterly*, *59*(1), 1–11. https://doi.org/10.1177/106591290605900101

Carneiro, R. L. (1970). A Theory of the Origin of the State: Traditional theories of state origins are considered and rejected in favor of a new ecological hypothesis. *Science*, *169*(3947), 733–738. https://doi.org/10.1126/science.169.3947.733

Chang, C. (2020). A data driven approach to study the social and political statuses of urban communities in kunming. *Journal of Chinese History*, *4*(2), 461–481. https://doi.org/10.1017/jch.2019.42

Chang, C., & Wang, Y. (2023). *Replication data for: The reach of the state*. Harvard Dataverse. Retrieved from.https://doi.org/10.7910/DVN/CSBRZX

Clarke, K., & Kocak, K. (2020). Launching revolution: Social media and the egyptian uprising's first movers. *British Journal of Political Science*, *50*(3), 1025–1045. https://doi.org/10.1017/s0007123418000194

Davenport, C. (2007). State repression and political order. *Annual Review of Political Science*, *10*, 1–23. https://doi.org/10.1146/annurev.polisci.10.101405.143216

Deng, Y., & O'Brien, K. J. (2013). Relational repression in China: Using social ties to demobilize protesters. *The China Quarterly*, *215*, 533–552. https://doi.org/10.1017/s0305741013000714

Dincecco, M. (2017). *State capacity and economic development*. Cambridge University Press.

Ekiert, G. (1991). Democratization processes in east central europe: A theoretical reconsideration. *British Journal of Political Science*, *21*(3), 285–313. https://doi.org/10.1017/s0007123400006177

Evans, P. B., Rueschemeyer, D., & Skocpol, T. (1985). *Bringing the state back in*. Cambridge University Press.

Fearon, J. D., & Laitin, D. D. (2003). Ethnicity, insurgency, and civil war. *American Political Science Review*, *97*(01), 75–90. https://doi.org/10.1017/s0003055403000534

Ferraro, J. M. (2003). *Family and public life in brescia, 1580–1650: The foundations of power in the Venetian state*. Cambridge University Press.

Francisco, R. A. (1996). Coercion and protest: An empirical test in two democratic states. *American Journal of Political Science*, *40*(4), 1179–1204. https://doi.org/10.2307/2111747

Friedman, D. (1977). A theory of the size and shape of nations. *Journal of Political Economy*, *85*(1), 59–77. https://doi.org/10.1086/260545

Fukuyama, F. (2004). The imperative of state-building. *Journal of Democracy*, *15*(2), 17–31. https://doi.org/10.1353/jod.2004.0026

Fukuyama, F. (2013). What is governance? *Governance*, *26*(3), 347–368. https://doi.org/10.1111/gove.12035

Gehlbach, S. (2008). What is a big bureaucracy? Reflections on rebuilding leviathan and runaway state-building. *Sociologickỳ časopis/Czech Sociological Review*, *44*(6), 1189–1198. https://doi.org/10.13060/00380288.2008.44.6.09

Goodwin, J. (1999). *No other way out*. Cambridge University Press.

Gunning, J., & Baron, I. Z. (2014). *Why occupy a square?* Oxford University Press.
Hanson, J. K., & Sigman, R. (2021). Leviathan's latent dimensions: Measuring state capacity for comparative political research. *The Journal of Politics*, *83*(4), 1495–1510. https://doi.org/10.1086/715066
Harbers, I. (2015). Taxation and the unequal reach of the state: Mapping state capacity in Ecuador. *Governance*, *28*(3), 373–391. https://doi.org/10.1111/gove.12117
Harbers, I., & Steele, A. (2020). Subnational variation across states: A typology and research agenda. *Latin American Politics and Society*, *62*(3), 1–18. https://doi.org/10.1017/lap.2020.4
Hendrix, C. S. (2010). Measuring state capacity: Theoretical and empirical implications for the study of civil conflict. *Journal of Peace Research*, *47*(3), 273–285. https://doi.org/10.1177/0022343310361838
Herbst, J. (2014). *States and power in africa*. Princeton University Press.
Hibbs, D. A. (1973). *Mass political violence: A cross-national causal analysis*. Wiley-Interscience.
Hollyer, J. R., Rosendorff, B. P., & Vreeland, J. R. (2015). Transparency, protest, and autocratic instability. *American Political Science Review*, *109*(4), 764–784. https://doi.org/10.1017/s0003055415000428
Hotelling, H. (1929). Stability in competition. *The Economic Journal*, *39*(153), 41–57. https://doi.org/10.2307/2224214
Huber, J. D., & Shipan, C. R. (2002). *Deliberate discretion? The institutional foundations of bureaucratic autonomy*. Cambridge University Press.
Kalyvas, S. N. (2006). *The logic of violence in civil war*. Cambridge University Press.
King, G., Honaker, J., Joseph, A., & Scheve, K. (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American Political Science Review*, *95*(1), 49–69. https://doi.org/10.1017/s0003055401000235
King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review*, *107*(2), 326–343. https://doi.org/10.1017/s0003055413000014
Koss, D. (2018). *Where the party rules*. Cambridge University Press.
Kuran, T. (1991). Now out of never: The element of surprise in the east european revolution of 1989. *World Politics*, *44*(1), 7–48. https://doi.org/10.2307/2010422
Lattimore, O. (1937). Origins of the great wall of China: A frontier concept in theory and practice. *Geographical Review*, *27*(4), 529–549. https://doi.org/10.2307/209853
Lee, M. M. (2020). *Crippling leviathan: How foreign subversion weakens the state*. Cornell University Press.
Lee, M. M., & Zhang, N. (2017). Legibility and the informational foundations of state capacity. *The Journal of Politics*, *79*(1), 118–132. https://doi.org/10.1086/688053
Levi, M. (1988). *Of rule and revenue*. University of California Press.
Levitsky, S. R., & Way, L. A. (2012). Beyond patronage: Violent struggle, ruling party cohesion, and authoritarian durability. *Perspectives on Politics*, *10*(4), 869–889. https://doi.org/10.1017/s1537592712002861

Lichbach, M. I. (1987). Deterrence or escalation? The puzzle of aggregate studies of repression and dissent. *Journal of Conflict Resolution*, *31*(2), 266–297. https://doi.org/10.1177/0022002787031002003

Lohmann, S. (1994). The dynamics of informational cascades: The monday demonstrations in leipzig, east Germany, 1989–91. *World Politics*, *47*(1), 42–101. https://doi.org/10.2307/2950679

Luna, J. P., & Soifer, H. D. (2017). Capturing sub-national variation in state capacity: A survey-based approach. *American Behavioral Scientist*, *61*(8), 887–907. https://doi.org/10.1177/0002764217720964

Malesky, E., Schuler, P., & Tran, A. (2012). The adverse effects of sunshine: A field experiment on legislative transparency in an authoritarian assembly. *American Political Science Review*, *106*(4), 762–786. https://doi.org/10.1017/s0003055412000408

Mann, M. (1984). The autonomous power of the state: Its origins, mechanisms and results. *European Journal of Sociology*, *25*(2), 185–213. https://doi.org/10.1017/s0003975600004239

Mattingly, D. C. (2016). Elite capture: How decentralization and informal institutions weaken property rights in China. *World Politics*, *68*(3), 383–412. https://doi.org/10.1017/s0043887116000083

Mattingly, D. C. (2017). Colonial legacies and state institutions in China: Evidence from a natural experiment. *Comparative Political Studies*, *50*(4), 434–463. https://doi.org/10.1177/0010414015600465

McAdam, D., & Su, Y. (2002). The war at home: Antiwar protests and congressional voting, 1965 to 1973. *American Sociological Review*, *67*(5), 696–721. https://doi.org/10.2307/3088914

McCarthy, J. D., McPhail, C., & Smith, J. (1996). Images of protest: Dimensions of selection bias in media coverage of Washington demonstrations, 1982 and 1991. *American Sociological Review*, *61*(3), 478–499. https://doi.org/10.2307/2096360

North, D. (1981). *Structure and change in economic history*. Norton.

North, D., & Thomas, R. (1973). *The rise of the western world: A new economic history*. Cambridge University Press.

O'Donnell, G. (1993). On the state, democratization and some conceptual problems: A Latin american view with glances at some postcommunist countries. *World Development*, *21*(8), 1355–1369. https://doi.org/10.1016/0305-750x(93)90048-e

Olson, M. (1982). *Rise and decline of nations: Economic growth, stagflation, and social rigidities*. Yale University Press.

Peic, G., & Reiter, D. (2011). Foreign-imposed regime change, state power and civil war onset, 1920–2004. *British Journal of Political Science*, *41*(3), 453–475. https://doi.org/10.1017/s0007123410000426

Pepinsky, T. B. (2019). The return of the single-country study. *Annual Review of Political Science*, *22*(1), 187–203. https://doi.org/10.1146/annurev-polisci-051017-113314

Pepinsky, T. B., Pierskalla, J. H., & Sacks, A. (2017). Bureaucracy and service delivery. *Annual Review of Political Science*, *20*(1), 249–268. https://doi.org/10.1146/annurev-polisci-051215-022705

Pop-Eleches, G., & Way, L. A. (2021). Censorship and the impact of repression on dissent. *American Journal of Political Science*, *67*(2), 456–471. https://doi.org/10.1111/ajps.12633

Rodden, J. (2006). *Hamilton's paradox: The promise and peril of fiscal federalism*. Cambridge University Press.

Rogowski, J. C., Gerring, J., Maguire, M., & Cojocaru, L. (2022). Public infrastructure and economic development: Evidence from postal systems. *American Journal of Political Science*, *66*(4), 885–901. https://doi.org/10.1111/ajps.12594

Rothstein, B. (2011). *The quality of government*. University of Chicago Press.

Rozenas, A., Schutte, S., & Zhukov, Y. (2017). The political legacy of violence: The long-term impact of stalin's repression in Ukraine. *The Journal of Politics*, *79*(4), 1147–1161. https://doi.org/10.1086/692964

Schelling, T. C. (1960). *The strategy of conflict*. Harvard University Press.

Scott, J. C. (1998). *Seeing like a state*. Yale University Press.

Scott, J. C. (2009). *The art of not being governed*. Yale University Press.

Shue, V. (1988). *The reach of the state*. Stanford University Press.

Skocpol, T., & Finegold, K. (1982). State capacity and economic intervention in the early new deal. *Political Science Quarterly*, *97*(2), 255–278. https://doi.org/10.2307/2149478

Slater, D. (2003). Iron cage in an iron fist: Authoritarian institutions and the personalization of power in Malaysia. *Comparative Politics*, *36*(1), 81–101. https://doi.org/10.2307/4150161

Soifer, H. D. (2008). State infrastructural power: Approaches to conceptualization and measurement. *Studies in Comparative International Development*, *43*(3–4), 231–251. https://doi.org/10.1007/s12116-008-9028-6

Soifer, H. D. (2015). *State building in Latin america*. Cambridge University Press.

Soifer, H. D. (2019). Units of analysis in subnational research. In A. Giraudy, E. Moncada, & R. Snyder (Eds.), *Inside countries: Subnational research in comparative politics* (pp. 92–112). Cambridge University Press.

Stasavage, D. (2010). When distance mattered: Geographic scale and the development of european representative assemblies. *American Political Science Review*, *104*(4), 625–643. https://doi.org/10.1017/s0003055410000444

Steinberg, J. (2018). 'strong' states and strategic governance: A model of territorial variation in state presence. *Journal of Theoretical Politics*, *30*(2), 224–245. https://doi.org/10.1177/0951629817748204

Stockmann, D. (2013). *Media commercialization and authoritarian rule in China*. Cambridge University Press.

Tarrow, S. (2018). Mann, war, and cyberspace: Dualities of infrastructural power in America. *Theory and Society*, *47*(1), 61–85. https://doi.org/10.1007/s11186-018-9306-x

Tilly, C. (1978). *From mobilization to revolution*. Addison-Wesley.

Tilly, C. (1992). *Coercion, capital and European states: AD 990 - 1992*. Wiley-Blackwell.

Touchton, M., Wampler, B., & Peixoto, T. (2021). Of democratic governance and revenue: Participatory institutions and tax generation in Brazil. *Governance*, *34*(4), 1193–1212. https://doi.org/10.1111/gove.12552

Truex, R. (2019). Focal points, dissident calendars, and preemptive repression. *Journal of Conflict Resolution*, *63*(4), 1032–1052. https://doi.org/10.1177/0022002718770520

Tsai, L. L. (2007). *Accountability without democracy: Solidary groups and public goods provision in rural China*. Cambridge University Press.

Tsai, L. L. (2008). Understanding the falsification of village income statistics. *The China Quarterly*, *196*, 805–826. https://doi.org/10.1017/s0305741008001136

Walker, C. F. (1999). *Smoldering ashes*. Duke University Press.

Wallace, J. L. (2016). Juking the stats? Authoritarian information problems in China. *British Journal of Political Science*, *46*(1), 11–29. https://doi.org/10.1017/s0007123414000106

Wang, Y. (2014). Empowering the police: How the Chinese communist party manages its coercive leaders. *The China Quarterly*, *219*, 625–648. https://doi.org/10.1017/s0305741014000769

Wang, Y. (2021). The political legacy of violence during China's cultural revolution. *British Journal of Political Science*, *51*(2), 463–487. https://doi.org/10.1017/s0007123419000255

Wang, Y., & Minzner, C. (2015). The rise of the Chinese security state. *The China Quarterly*, *222*, 339–359. https://doi.org/10.1017/s0305741015000430

Weber, M. (1946 [1918]). *Essays in sociology*. Oxford University Press.

Xu, Y., & Yao, Y. (2015). Informal institutions, collective action, and public investment in rural China. *American Political Science Review*, *109*(2), 371–391. https://doi.org/10.1017/S0003055415000155

Zhang, H., & Pan, J. (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. https://doi.org/10.1177/0081175019860244

Ziblatt, D. (2006). *Structuring the state*. Princeton University Press.

Author Biographies

Charles Chang is an Assistant Professor of Environment and Urban Studies at Duke Kunshan University.

Yuhua Wang is Professor in the Department of Government at Harvard University.