

Legibility and the Informational Foundations of State Capacity

Melissa M. Lee, Princeton University

Nan Zhang, Max Planck Institute for Research on Collective Goods

Recent research in political science has stressed the importance of the state in curbing violence and promoting social and economic development, resulting in an explosion of scholarly interest in the foundations of state capacity. This article argues that state capacity depends in part on “legibility”—the breadth and depth of the state’s knowledge about its citizens and their activities—and that legibility is crucial to effective, centralized governance. We illustrate the importance of legibility through a novel argument linking legibility to the state’s role in curbing free-riding in collective action dilemmas. We then demonstrate this argument in the context of tax contributions to public goods using an original measure of legibility based on national population censuses. The article concludes by discussing how future research may leverage our indicator’s exceptional temporal and geographic coverage to advance new avenues of inquiry in the study of the state.

In recent decades, political scientists have witnessed an explosion of interest in the state as a key actor in social and economic development. Indeed, variation in the effectiveness of state institutions has been linked to outcomes ranging from economic growth (Acemoglu and Robinson 2012; North 1981; Rothstein 2011), to civil war onset (Fearon and Laitin 2003; Hendrix 2010), to tax evasion and other forms of illicit activity (Brautigam, Fjeldstad, and Moore 2008). In addition, moving beyond a simple classification of “strong” and “weak” states, scholars and policy makers have begun to investigate the conceptual and empirical foundations of state capacity, with an eye toward improving the effectiveness of public institutions (Chong et al. 2014).

We contribute to this stream of scholarship by highlighting the importance of the state’s information gathering and processing functions. We argue that state capacity depends on the breadth and depth of the state’s knowledge about its citizens and their activities. Our informational account of state

capacity builds on the path-breaking work of James Scott, who held that a key component of governance involved making local practices “legible” to central state officials (Scott 1998, 2009). Legibility implies both (a) that the state possesses information about local practices and (b) that this information is rendered in standardized forms (e.g., cadastral maps, birth certificates, property registers) that are understandable to state administrators.

We illustrate the theoretical importance of legibility for state capacity by highlighting its role in facilitating the establishment of an efficient social order. Specifically, we argue that legibility is central to resolving the problem of free-riding in collective action settings as it allows the state to effectively monitor private behavior and enforce rules and regulations.¹ We apply this argument in the context of taxation and public goods provision, two key components of state capacity (Arbetman-Rabinowitz et al. 2012; Enriquez and Centeno 2012; Hanson and Sigman 2013; Hendrix 2010; Levi 1988;

Melissa Lee (melissa.lee@princeton.edu) is assistant professor of politics and international affairs in the Department of Politics and the Woodrow Wilson School for Public and International Affairs at Princeton University, Princeton, NJ 08544. Nan Zhang (zhang@coll.mpg.de) is senior research fellow at Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Strasse 10, D-53113, Bonn, Germany.

The authors gratefully acknowledge generous financial support from the National Science Foundation (no. DGE-0645962), the Freeman Spogli Institute Global Underdevelopment Action Fund, and Stanford University’s Center for African Studies. Data and supporting materials necessary to reproduce the numerical results in this article are available in the *JOP* Dataverse (<https://dataverse.harvard.edu/dataverse/jop>). An online appendix with supplementary material is available at <http://dx.doi.org/10.1086/688053>.

1. In this sense, our view of legibility is quite different from Scott’s original concept. While Scott sees the state as using legibility primarily to confiscate resources and enrich itself at the expense of the population, we stress that legibility is also central to the provision of social order and the control of free-riding.

Skocpol 1985). We therefore link legibility to the state's ability to curb free-riding in collective action dilemmas.

Empirically, we introduce a new operationalization of legibility based on the accuracy of age data reported in national population censuses. Our approach is robust to exogenous demographic shocks, is well correlated with other proxies of state capacity, and yields estimates of census error at both national and subnational levels across more than 120 countries for the period 1960–2012. We show that, controlling for economic, geographic, and temporal factors, areas with higher legibility also have more effective tax collection and enjoy greater public goods provision. Together the evidence supports our contention that legibility is a crucial, yet largely understudied, component of state capacity and administration.

Our research makes three key contributions to literature on the state and political development. First, we (re-)introduce researchers to the concept of legibility. We argue that failing to account for the state's informational functions omits a crucial variable in the study of state capacity, a central issue that occupies scholars and policy makers alike. Second, we advance the literatures on taxation and public goods provision by providing a novel argument linking legibility to the state's role in controlling opportunistic behavior in collective action settings. Finally, we introduce an original measure of legibility that can be calculated at both the national and subnational levels. Our indicator's exceptional temporal and geographic coverage has the potential to open up new avenues of research in the study of the state.

This article proceeds as follows. First, we describe the importance of legibility in greater detail, and we provide historical and contemporary examples of its centrality in resolving social dilemmas, focusing on the example of tax compliance. Following this discussion, we propose an operationalization strategy, describe our data and procedures, and present our estimates of age accuracy in the national census. We then demonstrate that census accuracy does indeed reflect a state's "presence on the ground," suggesting that our measure is a valid proxy for legibility more generally, and we provide additional validity checks by showing correlations with a number of perceptions-based indicators of state capacity. Finally, we explore the empirical relationships between legibility, tax contributions, and public goods outcomes. We conclude by discussing some implications of the present study for future research.

LEGIBILITY, COLLECTIVE GOODS, AND FREE-RIDING

"Legibility" refers to the breadth and depth of the state's knowledge of its citizens and their activities. At a basic level, the concept of legibility requires that the state collect in-

formation about the society it purports to govern. However, it is not enough that the state "sees" its citizens: the problem, as Scott (1998) insightfully argues, is that the kind of information that is useful to central authorities is not necessarily the same type of knowledge that is salient for private individuals. While individuals left to their own devices tend to develop complex, contextualized forms of knowledge suited to local social practices, these conventions are often incomprehensible to "outsiders" (i.e., state bureaucrats). Therefore, this local information must be organized into an administratively understandable format and thereby rendered "legible" to central state officials.

The contours of this challenge can be seen in the example of local weights and measures. In pre-modern Europe, units of measurement such as pints and *aunes* (a measure of length used for cloth) existed across many localities. However, the exact quantity denoted by these local units could vary dramatically. For instance:

The *pinte* in eighteenth-century Paris . . . was equivalent to 0.93 liters, whereas in Seine-en-Montagne it was 1.99 liters [and] in Precy-sous-Thil, an astounding 3.33 liters. The *aune* . . . varied depending on the material (the unit for silk, for instance, was smaller than that for linen), and across France there were at least seventeen different *aunes*. (Scott 1998, 25)

Given this welter of local standards, even if state authorities could know how many pints of beer or *aunes* of cotton cloth were produced in each locality, this knowledge would be extremely difficult to use for purposes of planning or comparison.

A second telling illustration comes from Scott's discussion of naming conventions in pre-modern Britain: "Tracking property ownership and inheritance, collecting taxes, maintaining court records, performing police work, conscripting soldiers, and controlling epidemics" all depended on the ability to identify individual subjects (Scott 1998, 71). However, this task was complicated by the fact that, outside of the aristocracy, most individuals did not use surnames, and 90% of the male population bore just six Christian names: John, William, Thomas, Robert, Richard, and Henry (Scott 1998, 68). To reduce confusion, a second designation could be added, indicating occupation (e.g., Smith, Baker), place of residence (e.g., Hill, Edgewood), father's given name (e.g., Johnson, O'Higgins), or a personal characteristic (e.g., Short, Strong). Although such naming conventions usually sufficed for local identification, they presented an administrative nightmare to central state agents who, naturally, were not privy to this local knowledge.

These two examples already hint at how the absence of legibility can facilitate opportunistic behavior in collective endeavors. Monitoring and enforcing compliance in such contexts requires that the state accurately identify individuals and measure how much each has contributed to the collective effort, such that it can know precisely who has undercontributed (and by how much). However, in the absence of administratively legible personal identities and units of account, free-riding is likely to be rampant.

We develop these ideas more fully in the context of contributions to public goods via taxes. This setting represents the quintessential example of a collective action dilemma: although society is better off if all comply with tax duties, each individual has an incentive to free-ride on the contributions of others. In this context, an important function of formal, centralized authority is to monitor opportunistic behavior and enforce fiscal rules, thereby sustaining cooperative outcomes (Greif 2006).² However, the state's ability to exercise these functions is compromised if social practices are illegible to central government officials.

The importance of legibility in preventing fiscal free-riding was already well known to early-modern state administrators. As the Marquis de Vauban explained to Louis XIV in 1686 when proposing an annual census: "Would it not be a great satisfaction to the king to know at a designated moment every year the number of his subjects, in total and by region, with all the resources, wealth and poverty of each place?" (Scott 1998, 11). In the absence of such detailed records, states developed imperfect work-arounds such as hearth taxes, or taxes on doors and windows, which solved the problem of knowing who should pay but which provided only rough approximations of the value in taxes that should be assessed.

These examples represent pre-modern versions of a modern day problem. Indeed, even in the age of regular incomes, pay slips, and W2 forms, it is still difficult for the state to evaluate exact tax burdens.³ Cash-only transactions, overseas earnings, and gambling winnings are all difficult to track, and the state must often rely upon self-declarations to learn anything about these sources. Indeed, the enduring problem of tax evasion in countries like Greece and Italy has much to do with the fact that large segments of the economy still operate

on a cash-only basis that leaves no paper trail visible to the state. Thus, while the contours of the information problem have changed in the modern world, its essential nature has not. Simply put, the problem of fiscal free-riding is in large part a problem of information.

However, even in the early-modern period, the legibility problem was not insurmountable. Consider the example of Sweden. On the surface, the Swedish state should have had great difficulty in preventing tax evasion because of the low degree of monetization of the economy and the weakness of the towns in generating commercial activity. However, these handicaps were offset by the success of the state in rendering legible the economic activities of both peasants and laborers, such that it could collect taxes from a broad population base (Tilly 1992, 136). The state's efforts to increase legibility dates to about 1540, when royal bailiffs were given the task of compiling detailed tax registers of peasant households (Hallenberg 2012). In 1628, the Swedish state undertook what some have called the first cadastral mapping in the Western world, which consisted of the production of around 12,000 maps delineating the boundaries of villages, freeholds, and farms, as well as details on land ownership, tenure, cultivation, and valuation (Hallenberg 2012; Kain and Baigent 1992). In addition, the Ecclesiastical Laws of 1686, which tasked the clergy with keeping lists of their parishioners' tax duties, further reinforced the state's capacity to gather fiscal information. The legacy of these efforts can be seen in the fact that, in the year 1920, already 80% of the economically active population in Sweden was registered with the tax authority (Flora and Heidenheimer 1981, 193).

The Swedish example also demonstrates that when society is made legible to the state, this can significantly curb the extent of fiscal free-riding. Such was the efficacy of monitoring and enforcement institutions that, between 1690 and 1873, the Swedish state could effectively levy taxes in kind on peasant households in order to support the army. Sweden was so successful in this regard that the state was able to keep 1.5% of its population under arms in the seventeenth century, which represented a higher rate of militarization than France, England, the Netherlands, or Russia (Tilly 1992, 59). As Steinmo (1996, 41) notes, "The hallmarks of the Swedish tax system have been its broad base, its stability, and its high yield," all of which depended fundamentally on a fiscally legible population.

OPERATIONALIZING LEGIBILITY

Having argued for the importance of legibility in the state's capacity to establish social order, in this section, we introduce our empirical strategy to operationalize legibility. Because we wish to adopt a comparative approach to measure

2. Micro-level evidence in support of this contention can be found in the cross-national experiments conducted by Herrmann, Thöni, and Gächter (2008), who find that cooperation is more robust in societies with stronger rule of law.

3. In this context, it is perhaps no surprise that "weak" states have traditionally relied upon customs and trade taxes, as more personal forms of income and wealth could be readily hidden (Brautigam et al. 2008).

legibility across political units and over time, we eschew more fine-grained measures of information and turn our attention instead to a coarser, but more broadly comparable, data series, the national census.

Historically the census has represented one of the state's primary sources of information on its population. Ancient Israel, ancient Rome, and Han Dynasty China conducted censuses for conscription and taxation purposes, and in modern times, states have used census information for purposes ranging from political apportionment and the allocation of funds to planning for the delivery of public goods and services. The census thus plays a central informational role in state administration.

We are, of course, not the first scholars to use the census in relation to the measurement of state capacity (Onorato, Scheve, and Stasavage 2014; Soifer 2013). However, earlier studies have typically operationalized the census in a blunt fashion, asking simply whether the state has conducted a census or not. This approach is less than ideal for two reasons. First, since only a small number of countries at the extreme end of the “failed states” spectrum do not collect census data, earlier studies miss substantial variation in the scope of the legibility problem. Second, this approach ignores subnational variation in census quality, which may be important and interesting in its own right.

Our legibility measure addresses these problems by considering the accuracy of information on exact ages within existing censuses. While there are, of course, many ways to gauge the accuracy of a census, we focus on information about exact ages for two reasons. First, age information is collected in almost all census enumerations, which increases the comparability of our measure. Second, it is relatively easy to quantify the scope of error in this data series based on existing demographic techniques.

In particular, it is well known among demographers that true age distributions within a population follow a naturally smooth curve. For example, as shown in the case of Switzer-

land in figure 1A, we see a clear underlying pattern to the distribution of ages, with small year-to-year changes in the count of individuals at each precise age. Further, while the shape of population curves may differ across societies (especially in relation to changing demographic patterns or cataclysmic demographic events such as war), even in these cases year-to-year changes in age counts still follow a roughly regular trend. To see this, consider figure 1B, which graphs population data from the French 2006 national census. The large plateau between ages 30 and 60 corresponds to the post-World War II baby boom and subsequent 1970s decline in fertility rates. However, even in response to these seismic demographic shifts, the age distribution itself remains smooth.

By contrast, when there is error in the age data, we observe discontinuous “heaping” on certain numbers that disrupts the smoothness of the underlying age curve. This is shown in the Sierra Leone census displayed in figure 1C. Since it is clearly implausible to believe that there would be a spike in births every five years, and that these spikes would exactly correspond to the ages 5, 10, 15, and so forth, we can attribute these patterns to errors in the data series. Below we introduce a method to quantify the magnitude of these errors. Before doing so, however, we first turn to a discussion of the data-generating process, as this touches upon the interpretation of our census measure.

Data-generating processes

Importantly, while we examine directly whether the state has accurate age information about the population, we also argue that census age errors proxy for the legibility problem more broadly. Specifically, the patterns we observe in figure 1C are likely to result (a) from a lack of age awareness in the population at large or (b) when census enumerators have difficulty finding or reaching the population to be counted. As we discuss below, one or both conditions are likely to obtain in “remote” regions where there is very little interaction between the state and society. Thus, when a state has poor

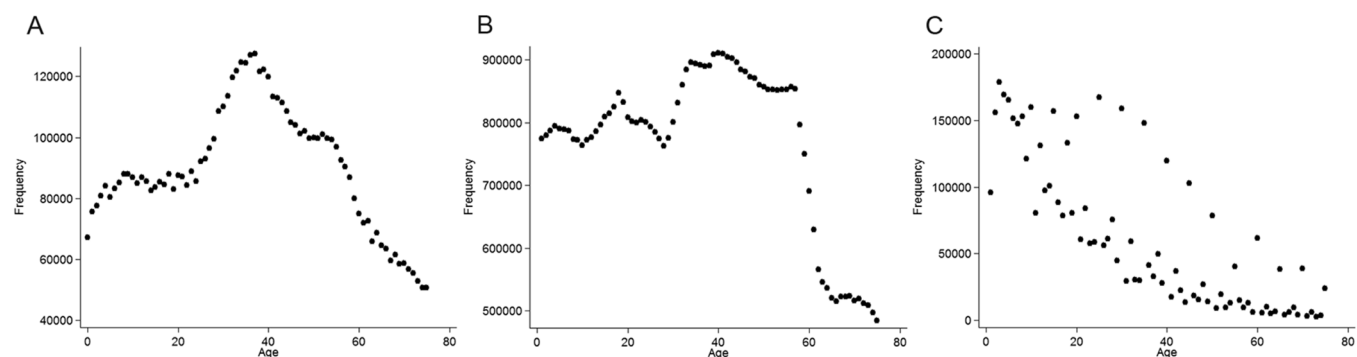


Figure 1. The effect of demographic shocks on the smoothness of age curves: A, Switzerland, 2000; B, France, 2006; C, Sierra Leone, 2004

information on the age of census respondents, this is likely to function as a “canary in a coal mine” indicator of poor legibility more generally.

With regard to the first mechanism, many scholars have documented that age heaping occurs when true ages are unknown (Duncan-Jones 2002; Herlihy and Klapisch-Zuber 1985; Nagi, Stockwell, and Snaveley 1973; Quandt 1973). In such cases, we might model the age recorded in the census as some guess that is near the true age. However, as both demographers and historians have noted, these guesses are generally not randomly distributed but rather tend to cluster or “heap” on certain numbers. The choice of focal numbers is different across societies and time periods, but it most often terminates with the digits “5” or “0” or (to a lesser extent) even numbers (Driscoll and Naidu 2012).

An alternative explanation for age heaping is that such patterns do not reflect ignorance about true ages but rather lying on the part of respondents. However, while individuals do lie to census enumerators (particularly with respect to sensitive topics), lying about one’s age tends to produce distinct patterns. For example, individuals may understate their age to avoid military conscription or women who bear children before a culturally appropriate age may overstate their age. As we demonstrate in the appendix (available online), if lying is present, we should observe heaping on only one focal number. However, we do not detect such “one-off” patterns in our data. Moreover, the above examples notwithstanding, age is generally apolitical and “low stakes” compared to other types of information that states might want to collect (such as property values, ethnicity, or geographic location), meaning that enumerators and respondents will have fewer reasons to intentionally misreport age information. Thus, we do not believe that age heaping is indicative of lying, but rather of general age unawareness.

More importantly for our purposes, where age awareness is low, legibility with respect to other issue domains is likely to be low as well. In fact:

A society in which individuals know their age only approximately is a society in which life is not governed by the calendar and the clock but by the seasonal cycle; in which birth dates are not recorded by families or authorities; in which few individuals must document their age in connection with privileges (voting, office-holding, marriage, holy orders) or obligations (military service, taxation). (A’Hearn, Baten, and Crayen 2009, 785)

There are two parts of this characterization worth stressing. First, it suggests that people learn their own exact ages

when the state provides incentives for them to do so. In fact, the state uses age as a critical piece of information to define eligibility for certain rights, responsibilities, and privileges.⁴ For example, in the contemporary world, one must be of age to vote, to serve in the military, to register for a driver’s license, to work legally, to enroll in public primary school, or to receive benefits like social security. As long as these rights and responsibilities are conditional on age, then the salience of age should increase with the density of state-society interactions, and so too will the likelihood that individuals learn and remember their precise ages.

Second, through interaction with state institutions, individuals obtain “artifacts” that help them to recall and track their ages as they grow older. For example, basic identification documents such as passports, birth certificates, and national ID cards often include a birthdate, allowing individuals to compare their year of birth with the current year to determine their ages. We can see the importance of these documents in the following exchange between a census enumerator and a respondent during the 1971 Moroccan census:

Enumerator: What is your age?

Respondent: Who me? Our generation was unrecorded. We didn’t have any. No date of birth. Nothing.

Enumerator: How many (years), how many? Estimate.

Respondent: How am I going to estimate? I have nothing to estimate with. I can tell you that I am 60 years; 70 I haven’t reached.

(Quandt 1973, 45; emphasis ours)

The issuing of these documents can also provide opportunities for individuals to learn their ages when they do not previously know them. For example, if a bureaucrat encounters an individual who does not know her age, the official can estimate her age by asking a series of probing questions that benchmark that person against national or local historical events or her relationship to other family members. In this way, accurate knowledge of the individual’s exact age may

4. Of course, many traditional societies attach importance to special ages (e.g., initiation ceremonies), but age cutoffs are sometimes indistinct (thereby grouping people into age cohorts instead of individual age categories), and age-specific rites often occur early in life, such that the importance of age diminishes over time, and true age is therefore forgotten as individuals grow older.

emerge.⁵ In addition, once that person possesses a document reflecting her true age, her information may then be accurately recorded by the state in all subsequent interactions.

In summary, the lack of age awareness is likely to obtain where the state generally has very little contact with its citizens. In such environments, citizens have few incentives to learn their true ages with any precision, and few opportunities to obtain documents that can help with age calculation and recall. These considerations imply that when age information is inaccurate or missing, other types of information about the population useful for state administration are likely to be missing as well, making census accuracy a proxy for legibility more broadly.

A second, parallel, data generating-process relates to the conduct of census enumerators. In particular, it is possible that individuals know their own true ages but enumerators do not record this information correctly. For example, instead of asking age-related questions directly to members of the household, enumerators may record ages based on second-hand information obtained from neighbors or local notables or else simply make up the data themselves. An account of challenges encountered during the 1961 census of Nepal acknowledged the problem of shirking by enumerators:

As the hill region was difficult to traverse, the enumerator would sit over an elevated place on a hill from where he could survey the surrounding settlements in the valleys and the ridges beyond. He would ask a local inhabitant about persons in the houses, which were visible from his place, and thus used this to collect population data of that area. (Kansakar 1977, 19)

Furthermore, because enumerators are drawn from the populations for which they are responsible, they tend to exhibit the same kind of digit biases as those held by respondents in their areas. These biases on the part of the enumerator also generate age heaping.⁶

While of course we cannot definitively gauge the severity of this problem, we note that systematic shirking by enumerators is likely under conditions of physical insecurity or inadequate infrastructure, which makes it dangerous or difficult to reach the population. For example, as Margot Anderson notes in the context of the nineteenth-century American frontier, “The U.S. population was extremely difficult to count not

only because it was primarily rural and spread over a huge area but also because decent local transportation often did not exist. The correspondence between officials in Washington, the federal marshals, and the enumerators is filled with tales of woe about reaching remote communities” (Anderson 1988, 25). Even into the 1870s, census enumeration in some parts of the American West “menaced by Indian attack, or frequented by lawless bands of Whites [required the] organization and equipment of a small expedition, including guides, interpreters, and even army escorts” (Anderson 1988, 90).

This dynamic reinforces our previous arguments with respect to the lack of age awareness because in such remote frontier areas there is likely to be sparse contact between the state and its citizens. Moreover, the absence of transportation infrastructure (almost always state provided) or the failure to maintain physical security also suggests little state-society interaction. Finally, if other state bureaucrats confront similar challenges as do census enumerators, they may likewise choose to shirk, thereby reducing the probability that citizens will transact with the state in other arenas as well. Thus, regardless of whether age heaping is the result of general age unawareness among the population or specific enumerator error, both data-generating processes are likely to indicate a broader absence of legibility.

Quantifying age accuracy

Demographers have developed several indices to quantify the extent of age heaping. These indices begin from the premise that, if there is no systematic irregularity in the reporting of true ages, then the distribution of the population by the terminal digits of their ages should be uniform. For instance, suppose we had population data containing reported ages between 15 and 74.⁷ If there is no heaping present, 10% of the population should report an age ending in 0, 10% ending in 1, and so on. The most straightforward index, the Whipple Index, simply calculates the percentage of the population with recorded ages ending in 5 or 0 and determines how much this percentage deviates from the expected value of 20%. The Whipple Index ranges from 100 (i.e., 20% / 20%), representing no preference for ages ending in 5 or 0, to 500 (i.e., 100% / 20%), representing a situation in which all reported ages have terminal digits of 5 or 0.

However, in natural populations, the assumption that the terminal digits distribute uniformly is violated by the effects of mortality. For instance, in our bin [15–74], we would expect fewer 74-year-olds than 73-year-olds, fewer 73-year-olds

5. See the appendix for an example from Indonesia.

6. It is unlikely that efforts to conceal shirking will succeed. For example, evidence using forensic techniques suggests that individuals are quite poor at hiding fraudulent data, even when the stakes are very high (Myagkov, Ordeshook, and Shakin 2009; Nigrini 1999).

7. The selection of the age bin is largely arbitrary as long as the number of ages in the bin is a multiple of 10.

than 72-year-olds, and so forth. The result is an overrepresentation of 5s, 6s, and 7s and an underrepresentation of 4s, 3s, and 2s. Moreover, this problem persists regardless of how we choose to define the age bin: there will always be an overrepresentation of the terminal digits at the beginning of the bin and an underrepresentation of the digits at the end.

To correct for this phenomenon, Myers (1940) developed a technique of creating a “blended” population that—assuming that true ages are correctly recorded—will return each terminal digit 10% of the time. Since beginning a bin at any given digit overstates the relative frequency of that digit, Myers’ technique does “complete justice” to each digit by starting at each one in turn. For example, the frequencies of each terminal digit are first tabulated for the bin [15–64] 10 times, starting from each terminal digit: first counting [15–64], then [16–65], continuing on until [24–73]. The 10 counts are summed and converted into a percentage of the grand population total. The resulting Myers Index ranges from 0, representing no heaping on any digit, to 90, representing the case where all ages were reported at a single terminal digit.⁸

We can return to our examples of Switzerland, France, and Sierra Leone and compute the Whipple score and Myers score for each distribution. Recall that the minimum values for Whipple and Myers are 100 and 0, respectively. Switzerland has a Whipple score of 102.86 and a Myers score of 0.22, indicating almost no error in the reported data. France has scores similar to Switzerland, even though its distribution is discontinuous due to war-related demographic shocks. These low scores corroborate what we saw graphically in figure 1. In contrast, we calculate a Whipple score of 228.20 and a Myers score of 26.48 for Sierra Leone in 2004. The much larger values on the indices indicate a high degree of age heaping, which we can confirm visually in figure 1C.

Researchers must confront a trade-off when choosing between the Whipple and Myers approaches. Myers is advantageous in that it does not require one to specify the precise digits on which heaping is likely occur, a property that is particularly attractive when working with cross-national or cross-cultural data where the exact form of digit preference (or avoidance) varies or is unknown. The Myers technique also accounts for mortality, while Whipple does not. However, the trade-off lies in scale invariance. The Myers Index is upwardly biased at very small population sizes (< 5,000 individuals) with very low true error rates (< 5%), while the Whipple Index performs much better under these conditions.⁹ For the analyses that follow, because our data are

drawn at the national level or from first-level subnational units (provinces, departments, or states) that tend to be sufficiently populous, we report Myers scores. Further, we restrict attention to developing countries, which are highly unlikely to have very low true error rates in their reported age data and therefore are less likely to be subject to the problem of scale variance.

Worldwide variation in legibility

Our full data set consists of 370 censuses covering the period 1960–2012.¹⁰ Nearly half of the observations represent original data that we collected directly from national census reports, making our data set on legibility the most comprehensive compendium of census-based data at the time of writing. Since most countries do not count their populations more frequently than once per decade, our data are not annual.¹¹ We do not regard this characteristic of our data as problematic, given that legibility likely changes slowly over time. Three features of the national-level Myers data stand out. First, we observe a considerable range in Myers scores for the whole world. The mean value in the sample is 8.21, which approximates a country like Côte d’Ivoire in 1998. Pakistan in 1973 has the worst score in the data set: its Myers Index is 45.67, meaning that 46% of the age data were misreported during that census year. On the opposite end of the distribution is Canada in 1991, with a score of 0.18. However, it would be misleading to characterize Canada as having greater legibility in 1991 than, say, Switzerland in 2011 (Myers = 0.23), as differences between Myers scores are not particularly meaningful below 1.00.

Second, the data exhibit a significant degree of skewness (fig. 2A). As with most political phenomena, this skewness can be attributed to between-country differences in legibility rather than to over-time differences. To see how this is so, consider figure 2B, which plots a histogram of the Myers scores after subtracting out the country means. After doing so, the data distribute approximately normally.

Third, looking across the world for the period 1960–2012 (table 1), Myers scores seem to reflect a plausible ranking of

8. The appendix describes how to calculate the Whipple and Myers Indices.

9. See the appendix for more on this point.

10. Census data were obtained via original data collection from national census reports or were downloaded from census data repositories (European Commission 2015; Latin American and Caribbean Demographic Center 2013; Minnesota Population Center 2013). The full list of censuses is reported in the appendix.

11. Some countries fail to conduct censuses on a regular basis. This failure is also informative: nearly all cases of which we are aware are cases where the country faced extreme limitations on state capacity. For example, war-ravaged Angola did not conduct its first postcolonial census until 2014, nearly 30 years after independence. Similarly, Somalia has not held a census since 1975.

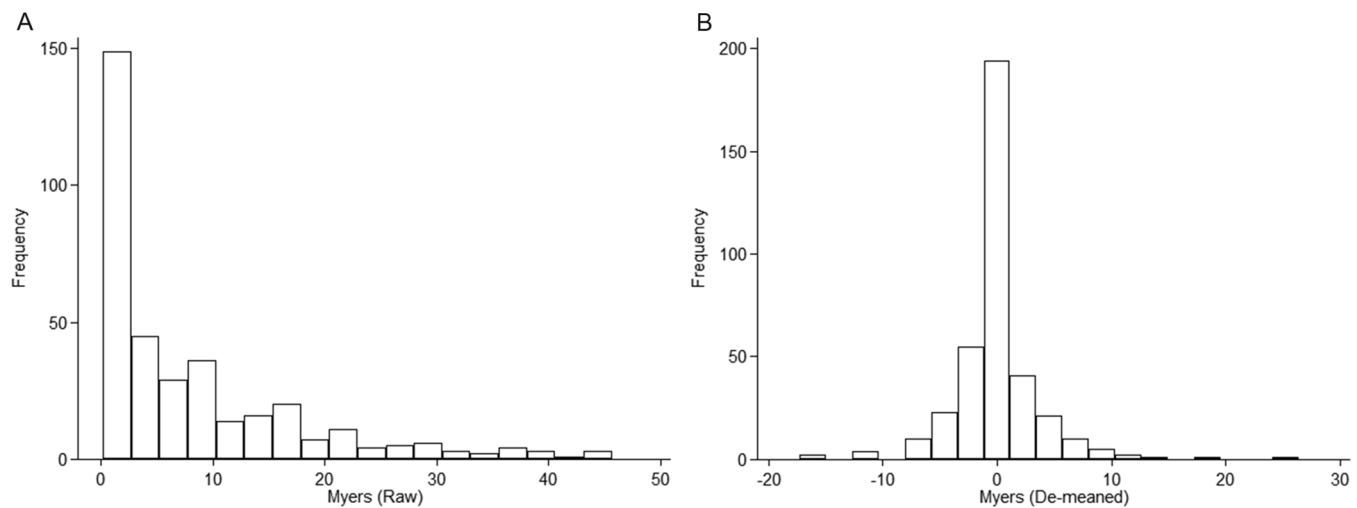


Figure 2. Histograms of national Myers scores: A, Raw Myers; B, De-meaned Myers

countries in terms of institutional effectiveness.¹² Asia, sub-Saharan Africa, and the Middle East/North Africa exhibit the widest range in Myers scores, with their maximum values (Pakistan 1971, Niger 1977, and Morocco 1960) close to or at the world maximum. Although all three regions have similar means, Asia's median is much smaller than the median for sub-Saharan Africa or the Middle East, suggesting that Asia has fewer extreme cases of poor legibility. We also see that Eastern Europe and Latin America look quite similar, though Eastern Europe has slightly smaller Myers scores. Finally, as one might expect, Western states enjoy very high levels of legibility. Figure 3 shows these patterns for the world for the 2005–14 round of censuses. In summary, countries that most analysts would intuitively consider to be high-capacity states have less age heaping.

VALIDATION

In this section, we demonstrate the plausibility of the Myers Index as a measure of legibility more broadly. In doing so, we draw the reader's attention back to our previous discussion of incentives and artifacts. More specifically, we argued that low age awareness is likely to prevail where the state has sparse contact with its citizens, such that it is likely to possess very little standardized information about local life. To substantiate these arguments, we examine the empirical relationship between Myers scores and two other measures of government information: registration with the state and birth certificates. As these data sources are available for only a select

number of countries, we restrict our attention to relationships at the subnational level.¹³ More information on these data is available in the appendix.

Birth registration is an important means through which the state obtains information about who is a citizen. By creating administrative records of individuals from the time of the birth, the state knows not only who was born when and where but also who is to be given access to rights, responsibilities, and privileges vis-à-vis the state. Indeed, birth registers were originally created for the purposes of taxation, conscription, and the management of property, and they only became significant for public health much later (Brumberg, Dozor, and Golombek 2012). Because birth registration and birth certificates serve as a means through which the state makes its population legible, we expect that the Myers Index will be correlated with registration and possession of a birth certificate. For ease of interpretation, we invert the Myers Index so that higher values indicate greater levels of legibility. The expected correlations with the Myers Index should be positive. The results in the upper panel of table 2 bear out this expectation.

As an additional validity test, we examine the national-level correlations between the Myers Index and common perception-based indices of state capacity. These indices are the International Country Risk Guide (ICRG), the Worldwide Governance Indicators (WGI), the Fragile States Index (FSI), and the Bertelsmann Transformation Index (BTI).¹⁴ If

13. Data source: DHS Program (2014).

14. Data sources: Bertelsmann Stiftung (2014), Fund for Peace (2014), Kaufmann, Kraay, and Mastruzzi (2014), and the PRS Group (2014). See the appendix.

12. We obtain similar results when calculating Myers scores at the subnational level.

Table 1. Summary Statistics, National Myers Scores, 1960–2012

Region	Observations	Mean	SD	Min	Median	Max
World	370	8.21	9.52	.18	3.70	45.67
Sub-Saharan Africa	109	13.14	9.22	1.03	10.51	44.97
Asia	79	10.10	12.49	.83	3.40	45.67
Eastern Europe/former USSR	21	1.51	.85	.41	1.28	3.27
Latin America	85	4.98	4.53	.50	3.20	17.36
Middle East/North Africa	25	12.15	10.25	1.10	9.69	43.97
Western democracies/Japan	51	.98	.69	.18	.76	3.18

Sources. Regions are from the Minorities at Risk Project (2009).

we are correct that legibility is related to state capacity, we should observe a positive relationship between the (inverted) Myers Index and all indicators, with the exception of the FSI, which is scaled so that higher values indicate worse outcomes. Indeed, as the lower panel of table 2 shows, the correlations are relatively strong and signed in the predicted direction.

RESULTS: LEGIBILITY, TAXATION, AND COLLECTIVE GOODS

Next we turn to our investigation of how legibility helps the state promote an efficient social order. As we have argued, access to administratively useful information allows the state to monitor individual behavior and enforce compliance with rules and regulations. We test two implications of these claims in the context of contributions to public goods via taxes: greater legibility is associated with (a) higher tax compliance

and (b) better outcomes related to the delivery of public goods and services. In doing so, we demonstrate not only the empirical contributions of the Myers Index but also the theoretical and substantive importance of the concept of legibility.

To assess the relationship between legibility and taxation, we exploit subnational variation within countries. Our sample contains 12 countries: Argentina, Brazil, Greece, India, Indonesia, Italy, Mexico, Philippines, South Africa, Tanzania, Thailand, and Turkey. Our rather restricted sample reflects the fact that province-level data on tax revenues and gross domestic product (GDP) are extraordinarily difficult to obtain. However, the subnational coverage that comes at the expense of cross-national coverage is an asset: a subnational analysis will allow us to include country-level fixed effects to control for unmeasured differences across fiscal regimes (e.g., the tax rate) that can affect the final amount of taxes collected.

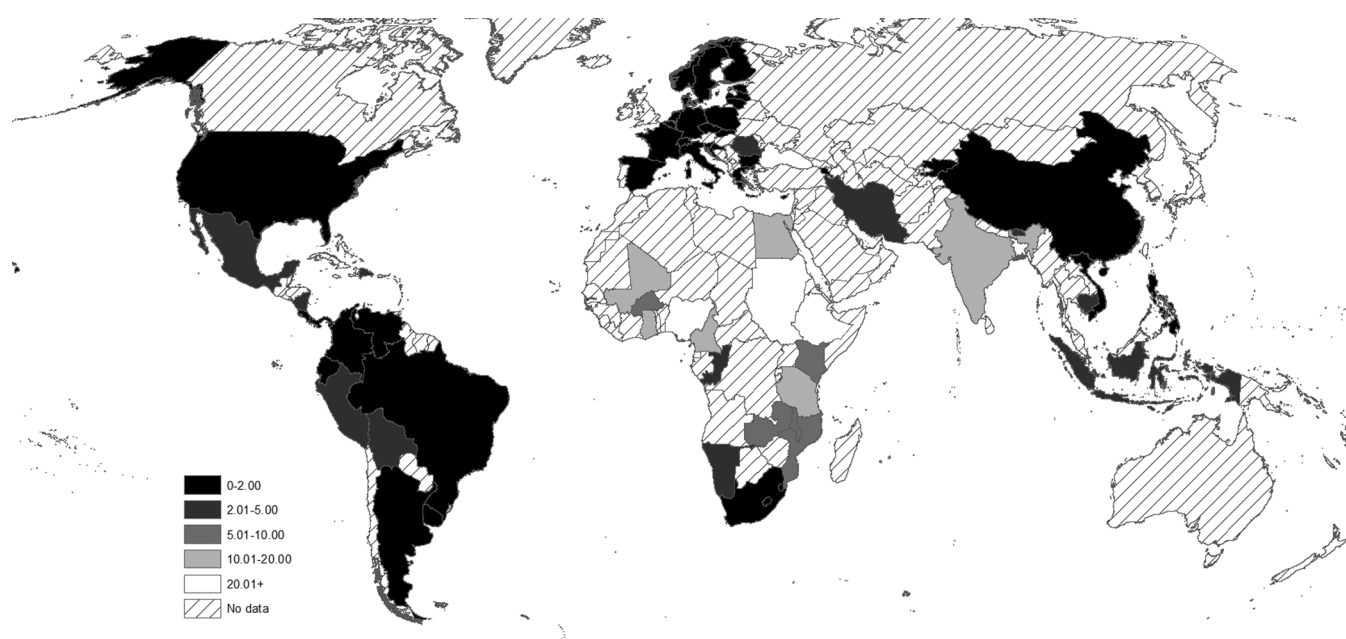


Figure 3. Myers scores by country, 2005–12

Table 2. Validity Checks for the Myers Index

Indicator	Observations	Coverage	Correlation	
			Raw Myers	Logged Myers
Subnational correlations:				
Birth registration	393	2000–2012	.444	.519
Birth certificate	282	2000–2012	.336	.411
National correlations:				
ICRG:				
Index total	232	1980–2012	.484	.610
Internal conflict	225	1980–2012	.461	.565
Bureaucratic quality	225	1980–2012	.436	.581
WGI:				
Government effectiveness	201	1990–2012	.476	.649
Political stability	202	1990–2012	.480	.603
Rule of law	202	1990–2012	.438	.627
Regulatory quality	202	1990–2012	.496	.616
Control of corruption	201	1990–2012	.442	.623
FSI:				
Index total	129	2000–2012	−.490	−.683
Public services	129	2000–2012	−.490	−.696
Security	129	2000–2012	−.417	−.559
BTI:				
Index total	99	2000–2012	.448	.550
Stateness	99	2000–2012	.495	.515

Note. ICRG = International Country Risk Guide; WGI = Worldwide Governance Indicators; FSI = Fragile States Index; BTI = Bertelsmann Transformation Index. Myers scores are inverted so that higher values indicate greater legibility. All correlations have the predicted sign and are statistically significant at $p < .01$.

We examine two dependent variables: *tax revenue*, defined as the amount of income tax collected by province, and *tax ratio*, or the ratio of income tax collected to province-level GDP.¹⁵ While the former presents a raw measure of fiscal contributions, the latter can be interpreted as a measure of the efficiency of the state in collecting taxes. We also control for factors likely to affect taxation at the subnational level: for example, states may find it more challenging to assess taxes from provinces that are *distant* from the capital, whereas high *population densities* might promote economies of scale in tax collection and enforcement. Finally, we include a measure of *regional GDP* in the tax revenue model to control for the fact that in more developed areas, there is simply more economic output to tax. All variables have been log-transformed due to skewness. Descriptive statistics are given in table 3.

We test the effect of legibility on taxation using an ordinary least squares (OLS) model, with standard errors clus-

tered by country and Myers scores lagged by one year.¹⁶ For ease of interpretation, we standardize our variables to have mean 0 and standard deviation 1. We also invert the Myers Index such that higher values indicate greater legibility. As shown in table 4, the coefficient on Myers is positive and statistically significant in all four models, as predicted by our theory. In the model that controls for likely confounds, we find that a one standard deviation shift in Myers scores results in about a 10% shift in tax revenues. This shift is consequential. Consider the example of India in 2012. Uttar Pradesh is among the poorest Indian states in per capita terms. In 2012, regional GDP stood at US\$372 per person, and its Myers score locates it in the bottom quartile in terms of legibility. The Indian state collected about \$2.3 billion in income tax from Uttar Pradesh in 2012. A shift of 10% of a

15. A full discussion of all measures in this section appears in the appendix.

16. This specification decision allows us to rule out the possibility of reverse causality. It also allows us to maximize observations. The substantive conclusions stand if we use contemporaneous Myers scores, but the estimates are less precise due to the loss of three countries and more than 100 observations. See the appendix.

Table 3. Summary Statistics: Taxation Analysis

Variable	Mean	SD	Min	Median	Max
Tax revenue					
per capita	4.62	3.41	−5.98	3.71	12.20
Tax ratio	−4.82	6.73	−21.36	−3.81	3.53
Legibility	−1.11	.95	−3.03	−1.07	1.32
Regional GDP					
per capita	9.45	5.15	.54	7.66	21.79
Distance	6.32	1.04	−.26	6.46	8.17
Population density	4.13	1.64	−.59	3.99	10.87
Terrain roughness	.46	1.14	−4.47	.78	2.43

Note. Number of observations = 399. All variables are logged. Myers scores are inverted.

standard deviation in legibility for Uttar Pradesh represents an increase of nearly \$320 million in additional income tax revenue—a very large sum for a single subnational unit.

We find similarly powerful effects of legibility on the tax ratio. As seen in column 4 of table 4, a one standard deviation increase in legibility results in a 6% standard deviation increase in the tax ratio. To use the example of India again, this shift would result in an increase of Uttar Pradesh's tax ratio from about 3% to about 6.5%. Thus, we observe that greater legibility is associated with better tax collection, and this effect is both substantively and statistically significant.¹⁷

We now demonstrate that legibility is also associated with better public goods provision. Since we cannot test service delivery directly, we examine outcomes related to service delivery: *infant mortality* as well as *adult literacy* and *primary school enrollment*. Here we leverage cross-national variation to maximize geographic and temporal coverage. In addition, since we consider health and education outcomes that are nearly universally desirable, we are less concerned that variation in these outcomes is driven by unobserved “demand effects” necessitating the use of country fixed effects. Our sample covers up to 111 countries from all regions of the world. Descriptive statistics are given in table 5.

Infant mortality is defined as the number of deaths in the first year of life, and we log the data to account for skewness. *Adult literacy* and *primary school enrollment* are both expressed as percentages. We also control for several covariates: *logged GDP per capita*, *regime type* operationalized using the Polity index, *logged population density*, and *logged terrain ruggedness*. As before, we standardize the variables to have

17. Our results are robust to bootstrapping the standard errors (see the appendix).

mean 0 and standard deviation 1, and we invert the Myers score so that higher values indicate greater legibility. Due to the scarcity of public goods data prior to 2000, we do not always have data on our dependent variables in the same year we obtain a Myers score. For that reason, all variables represent decade averages. Because legibility, like state capacity more generally, changes slowly over time, decade averages should not substantially alter our conclusions except through the introduction of measurement error. More information about the variables and sample can be found in the appendix.

We examine the effect of legibility on public goods outcomes with a simple OLS regression model that includes decade fixed effects and country-clustered standard errors. Table 6 presents the results for the three dependent variables. Looking at columns 2, 4, and 6, which show the results with the full set of covariates, we see that legibility is statistically significant and signed according to prediction. Greater legibility is associated with lower infant mortality rates and higher literacy and primary school enrollment rates. The

Table 4. Legibility and Taxation: Subnational Results

	Tax Revenue (1)	Tax Revenue (2)	Tax Ratio (3)	Tax Ratio (4)
Legibility	.319* (.138)	.104* (.0385)	.0634* (.0249)	.0587* (.0225)
Regional GDP per capita		1.582 ** (.0846)		
Distance		−.0230+ (.0124)		−.0161+ (.00753)
Population density		.0549+ (.0269)		.0216+ (.0107)
Terrain ruggedness		.0159 (.0161)		.00697 (.00813)
Constant	−.0817* (.0356)	−.0429** (.0102)	−.0195* (.00639)	−.0159* (.00597)
R ²	.0825	.781	.0503	.121

Note. Number of observations = 399. Number of countries = 12. Myers scores are inverted so that higher values indicate greater legibility. Country-specific intercepts suppressed. Standard errors are in parentheses and are clustered by country.

+ $p < .10$.

* $p < .05$.

** $p < .01$.

Table 5. Summary Statistics: Public Goods Analysis

Variable	Observations	Mean	SD	Min	Median	Max
Infant mortality rate	326	3.51	1.14	.51	3.80	5.24
Adult literacy rate	188	72.66	24.87	9.43	81.92	99.80
Primary school enrollment rate	244	81.63	19.83	13.66	89.33	99.84
Legibility	329	-1.39	1.28	-3.82	-1.26	1.70
GDP per capita	329	8.28	1.24	5.81	8.23	11.50
Democracy	329	2.37	6.68	-10.00	4.90	10.00
Population density	329	-10.23	1.41	-14.66	-10.10	-4.89
Terrain ruggedness	329	.36	.95	-2.41	.48	2.27

Note. Infant mortality, legibility, GDP per capita, population, and terrain are logged. The Myers Index is inverted.

relationship is substantively large in all three cases.¹⁸ For Kenya in the 2000s, for example, a one standard deviation increase in legibility is associated with an effect of about 17 fewer infant deaths per 1,000 births and 11 and 3 percentage point increases in the literacy and primary enrollment rates, respectively.¹⁹

Taken together, the results of our subnational and cross-national exercises are consistent with the claim that legibility is central to the state's capacity to curb free-riding in collective endeavors. We caution our readers that our results are suggestive and should not be interpreted as causal.²⁰ Even so, two takeaways are worth mentioning. First, if legibility played no role in helping the state promote and sustain good collective outcomes, we would not observe significant relationships between legibility, taxation, and public goods, particularly in the presence of powerful explanatory factors such as GDP per capita and regime type. For this reason, these results serve as an additional validity check on the Myers Index as a proxy for legibility and state capacity more generally. Second, our application in the domain of taxation and public goods demonstrates the potential for the concept of legibility and our census-based measure to shed new light on important theoretical and empirical issues of interest to a broad range of scholars. As we show, the geographic cov-

erage of the Myers Index allows us to explore variation at both the subnational and cross-national levels.

CONCLUSION

In this article, we highlight the importance of legibility—defined as the breadth and depth of the state's knowledge of its citizens and their activities—as a crucial but often neglected factor for understanding state capacity. We argue that legibility provides a key foundation for social and economic development by facilitating the centralized monitoring of free-riding in collective action settings. We illustrate this argument in the context of contributions to public goods via taxes using an original measure of legibility based on census data. In doing so, we highlight the importance of the state as an external monitor and enforcer in controlling opportunistic behavior and supporting an efficient social order.

While we believe that these theoretical and empirical contributions are important in their own right, we emphasize that our focus on legibility also has implications for the state capacity literature more broadly. Our tax compliance example illustrates that legibility is associated with both quantitative and qualitative changes in the nature of state power. At a basic level, the collection of existing fiscal dues is much more effective in areas where legibility is greater. However, this ability to read society also allows the state to deploy more comprehensive fiscal instruments (e.g., income taxes) than would otherwise be possible. Indeed, the movement from indirect to direct forms of taxation parallels the development of better sources of information on citizens and their economic activities, thereby making society more legible to central tax collectors (Jones 1988; Kiser and Sacks 2009; Martin, Mehrotra, and Prasad 2009). In other words, taking legibility into account can help to explain not only a quantitative increase in the state's extractive capacity but also a qualitative shift in the arsenal of extractive instruments at the state's disposal.

18. These results hold if we lag the Myers Index to account for the possibility that past levels of legibility may influence school-related outcomes. See the appendix.

19. As before, our results are robust to bootstrapping the standard errors.

20. An additional word of caution is in order: measures of tax revenue, tax ratio, literacy, and enrollment are plagued with missing data, measurement error, and even misrepresentation by government authorities. However, these problems are most likely to occur in weaker states, and they can therefore be expected to temper, rather than augment, the positive relationship between legibility and the variables we examine. As such, to the extent that bias exists, it would tend to work against our ability to find statistically significant effects.

Table 6. Legibility and Public Goods: National-Level Results

	Mortality (1)	Mortality (2)	Literacy (3)	Literacy (4)	Enrollment (5)	Enrollment (6)
Legibility	-.663** (.0398)	-.283** (.0481)	.797** (.0649)	.507** (.0839)	.586** (.0867)	.229* (.0942)
GDP per capita		-.531** (.0572)		.355** (.0738)		.367** (.0955)
Democracy		-.0835* (.0364)		.104 (.0726)		.0736 (.0636)
Population density		-.146** (.0322)		.0374 (.0456)		.131 ⁺ (.0678)
Terrain ruggedness		.0551 ⁺ (.0281)		.113 ⁺ (.0581)		.184** (.0655)
Constant	.333** (.0687)	.221** (.0531)	-.597** (.212)	-.514** (.193)	-.426* (.163)	-.413** (.147)
Number of observations	326	326	188	188	244	244
Number of countries	111	111	84	84	105	105
R ²	.744	.888	.673	.758	.445	.576

Note. Decade-specific intercepts are suppressed. Standard errors are in parentheses and are clustered by country.

⁺ $p < .10$.

* $p < .05$.

** $p < .01$.

Perhaps the most important empirical contribution of this study involves the production and dissemination of our legibility data set. We wish to highlight in particular the geographic and temporal coverage of our measure, which can be calculated for any national or subnational unit where single-age data are available. As we demonstrate in this article, this feature permits a level of flexibility in specifying the unit of analysis that will likely interest both comparativists and international relations scholars. Additionally, the Myers Index can be calculated for historical and future censuses as well. Indeed, to our knowledge, no existing indicator of state capacity has the cross-national, subnational, or historical coverage of our measure, and our ongoing data collection efforts promise to expand the coverage of the data set even further. We also stress that, unlike existing state capacity indicators, a key feature of our measure is its sensitivity to variation at the middle ranges of state capacity rather than at the extremes. Since most states are neither fully developed nor fully “failed,” the majority of the world’s countries lie in this middle range.

Our theoretical insights and data have implications for a range of research programs. Consider the literature on intrastate conflict. Although the study of civil war has been an especially productive line of research, scholars continue to disagree about the role of state capacity in explaining civil war onset, termination, or conflict processes. Part of this problem stems from the difficulty in specifying conceptually clear

definitions of state capacity and the inability to arbitrate between competing state capacity mechanisms. For instance, Fearon and Laitin (2003) suggested an information story in their discussion of insurgency, but they were unable to test this mechanism directly. Empirically, our approach provides one means for scholars to unpack state capacity into its constituent parts and to subject different explanations to more rigorous empirical scrutiny. In addition, at a theoretical level, our work has implications for explanations of conflict onset. If legibility is crucial for both the extraction of resources and the provision of public goods, states that are able to make their populations legible are likely to be the same states that can reduce the risk of civil war by providing the kinds of goods and services to populations that might otherwise support rebel organizations. Such states are also likely to be able to amass the fiscal resources necessary for funding their coercive apparatuses.

Our study also has theoretical implications for the literature on political and economic development. In the Hobbesian tradition, the regulation of private (mis)behavior and the resolution of collective action problems constitute the *raison d’être* of the state. Yet the existing literature has not adequately explored how exactly the state undertakes these tasks and why some states perform better in this role than others. Our project sheds light on this puzzle by highlighting legibility as a central ingredient of state capacity. Since legibility is crucial for de-

tering free-riding, states that possess more legible information about society should be better able to exercise key monitoring and enforcement functions. Our research on legibility thus illuminates an important stepping-stone in institutional development.

Finally, our data can also open up novel avenues for research on state capacity. For example, scholars could treat legibility itself as a dependent variable and investigate how states go about increasing the information at their disposal.²¹ Our discussion suggests that incentives and artifacts play important roles in this regard, but attempts to alter these parameters may backfire (Scott 1998, 2009). Indeed, Scott's work not only focuses on the importance of legibility for state action but also highlights how efforts to increase legibility were resisted by society. This suggests that the state's legitimacy is likely to be crucial to the development of legibility, and this opens the door for new linkages between the state capacity and legitimacy literatures. It also suggests that improving legibility is both a technical problem and a deep political challenge, as evidenced by contemporary debates in the United States about privacy and surveillance.

In summary, we are convinced that a renewed focus on legibility has the potential to yield novel theoretical and empirical insights into a range of outcomes of interest to political scientists. We believe that the present article has only begun to explore these exciting new research frontiers.

ACKNOWLEDGMENTS

We thank Jim Fearon, Frank Fukuyama, Judy Goldstein, Steve Krasner, David Laitin, Brian Min, Ken Scheve, Ken Schultz, and Hillel Soifer for their helpful comments. We also thank Héctor González Medina, Jasmine Dehghan, Daniel Flores, Jennie Lummis, Louis McWilliams, Alya Naqvi, Sanjana Parikh, and Peter Pham for research assistance.

REFERENCES

- Acemoglu, Daron, and James Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. New York: Random House.
- A'Hearn, Brian, Jörg Baten, and Dorothee Crayen. 2009. "Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital." *Journal of Economic History* 69 (3): 783–808.
- Anderson, Margo J. 1988. *The American Census: A Social History*. New Haven, CT: Yale University Press.
- Arbetman-Rabinowitz, Marina, Jacek Kugler, Mark Abdollahian, Kang Kyungkook, Hal T. Nelson, and Ronald L. Tammen. 2012. "Political Performance." In Jacek Kugler and Ronald L. Tammen, eds., *The Performance of Nations*. Lanham, MD: Rowman & Littlefield, 7–47.
- Bertelsmann Stiftung. 2014. *The Bertelsmann Stiftung's Transformation Index*. Gutersloh: Bertelsmann Stiftung.
- Besley, Timothy, and Torsten Persson. 2010. "State Capacity, Conflict, and Development." *Econometrica* 78 (1): 1–34.
- Brautigam, Deborah, Odd-Helge Fjeldstad, and Mick Moore, eds. 2008. *Taxation and State-Building in Developing Countries: Capacity and Consent*. Cambridge: Cambridge University Press.
- Brumberg, H. L., D. Dozor, and S. G. Golombek. 2012. "History of the Birth Certificate: From Inception to the Future of Electronic Data." *Journal of Perinatology* 32 (6): 407–11.
- Chong, Alberto, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer. 2014. "Letter Grading Government Efficiency." *Journal of the European Economic Association* 12 (2): 277–99.
- DHS Program. 2014. *Demographic and Health Surveys*. Rockville, MD: ICF International.
- Driscoll, Jesse, and Suresh Naidu. 2012. "State-Building and Census-Taking: The Political Economy of Population Data." Working paper, University of California, San Diego.
- Duncan-Jones, Richard. 2002. *Structure and Scale in the Roman Economy*. Cambridge: Cambridge University Press.
- Enriquez, Elaine, and Miguel Angel Centeno. 2012. "State Capacity: Utilization, Durability, and the Role of Wealth versus History." *Interdisciplinary and Multidisciplinary Journal of Social Sciences* 1 (2): 130–62.
- European Commission. 2015. *Eurostat*. Luxembourg: Eurostat. <http://ec.europa.eu/eurostat/web/main>.
- Fearon, James D., and David D. Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97 (1): 75–90.
- Flora, Peter, and Arnold Heidenheimer. 1981. *The Development of Welfare States in Europe and America*. New Brunswick: Transaction.
- Fund for Peace. 2014. *Fragile States Index*. Washington, DC: Fund for Peace.
- Greif, Avner. 2006. *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*. Cambridge: Cambridge University Press.
- Hallenberg, Mats. 2012. "For the Wealth of the Realm: The Transformation of the Public Sphere in Swedish Politics, c. 1434–1650." *Scandinavian Journal of History* 37 (5): 557–77.
- Hanson, Jonathan K., and Rachel Sigman. 2013. "Measuring State Capacity for Comparative Political Research." Working paper, Syracuse University.
- Hendrix, Cullen S. 2010. "Measuring State Capacity: Theoretical and Empirical Implications for the Study of Civil Conflict." *Journal of Peace Research* 47 (3): 273–85.
- Herlihy, David, and Christiane Klapisch-Zuber. 1985. *Tuscans and Their Families: A Study of the Florentine Catasto of 1427*. New Haven, CT: Yale University Press.
- Herrmann, Benedikt, Christian Thöni, and Simon Gächter. 2008. "Anti-social Punishment across Societies." *Science* 319 (5868): 1362–67.
- Jones, Carolyn. 1988. "Class Tax to Mass Tax: The Role of Propaganda in the Expansion of the Income Tax during World War II." *Buffalo Law Review* 37:685–738.
- Kain, Roger, and Elizabeth Baigent. 1992. *The Cadastral Map in the Service of the State: A History of Property Mapping*. Chicago: University of Chicago Press.
- Kansakar, Vidya Bir Singh. 1977. *Population Censuses of Nepal and the Problems of Data Analysis*. Kathmandu: Center for Economic Development and Administration.
- Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2014. *Worldwide Governance Indicators*. Washington, DC: World Bank.
- Kiser, Edgar, and Audrey Sacks. 2009. "Improving Tax Administration in Contemporary African States: Lessons from History." In Isaac Martin, Ajay Mehrotra, and Monica Prasad, eds., *The New Fiscal Sociology: Taxation in Comparative and Historical Perspective*. Cambridge: Cambridge University Press, 183–200.

21. For a formal theoretical treatment of "investments" in state capacity more broadly, see Besley and Persson (2010).

- Latin American and Caribbean Demographic Center. 2013. *REDATAM 7*. Santiago, Chile: Economic Commission for Latin America and the Caribbean. <http://celade.cepal.org/redbin/RpWebEngine.exe/Portal?lang=eng> (accessed October 3, 2016).
- Levi, Margaret. 1988. *Of Rule and Revenue*. Berkeley: University of California Press.
- Martin, Isaac William, Ajay Mehrotra, and Monica Prasad. 2009. *The New Fiscal Sociology: Taxation in Comparative and Historical Perspective*. Cambridge: Cambridge University Press.
- Minnesota Population Center. 2013. *Integrated Public Use Microdata Series, International: Version 6.2*. Minneapolis: University of Minnesota.
- Minorities at Risk Project. 2009. *Minorities at Risk Dataset*. College Park, MD: Center for International Development and Conflict Management.
- Myagkov, Mikhail, Peter C. Ordeshook, and Dimitri Shakin. 2009. *The Forensics of Election Fraud: Russia and Ukraine*. Cambridge: Cambridge University Press.
- Myers, Robert J. 1940. "Errors and Bias in the Reporting of Ages in Census Data." *Transactions of the Actuarial Society of America* 41 (104): 394–415.
- Nagi, Mostafa H., Edward G. Stockwell, and L. M. Snively. 1973. "Digit Preference and Avoidance in the Age Statistics of Some Recent African Censuses: Some Patterns and Correlates." *International Statistical Review* 41 (2): 165–74.
- Nigrini, Mark J. 1999. "I've Got Your Number." *Journal of Accountancy* 187 (5): 79–83.
- North, Douglass C. 1981. *Structure and Change in Economic History*. New York: Norton.
- Onorato, Massimiliano Gaetano, Kenneth Scheve, and David Stasavage. 2014. "Technology and the Era of the Mass Army." *Journal of Economic History* 74 (2): 449–81.
- PRS Group. 2014. *International Country Risk Guide*. East Syracuse, NY: PRS Group.
- Quandt, Anna Spitzer. 1973. "The Social Production of Census Data: Interviews from the 1971 Moroccan Census." PhD diss., University of California, Los Angeles.
- Rothstein, Bo. 2011. *The Quality of Government: Corruption, Social Trust, and Inequality in International Perspective*. Chicago: University of Chicago Press.
- Scott, James C. 1998. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven, CT: Yale University Press.
- Scott, James C. 2009. *The Art of Not Being Governed: An Anarchist History of Upland Southeast Asia*. New Haven, CT: Yale University Press.
- Skocpol, Theda. 1985. "Bringing the State Back In: Strategies of Analysis in Current Research." In Peter Evans, Dietrich Rueschemeyer, and Theda Skocpol, eds., *Bringing the State Back In*. Cambridge: Cambridge University Press, 3–37.
- Soifer, Hillel David. 2013. "Regionalism and State Weakness." Presented at the Instituto de Ciencia Política, Pontificia Universidad Católica de Chile, Santiago.
- Steinmo, Sven. 1996. *Taxation and Democracy: Swedish, British, and American Approaches to Financing the Modern State*. New Haven, CT: Yale University Press.
- Tilly, Charles. 1992. *Coercion, Capital, and European States, AD 990–1992*. Malden, MA: Wiley-Blackwell.