Commentary

What Is Governance?

FRANCIS FUKUYAMA*

This commentary points to the poor state of empirical measures of the quality of states, that is, executive branches and their bureaucracies. Much of the problem is conceptual, as there is very little agreement on what constitutes high-quality government. The commentary suggests four approaches: (1) procedural measures, such as the Weberian criteria of bureaucratic modernity; (2) capacity measures, which include both resources and degree of professionalization; (3) output measures; and (4) measures of bureaucratic autonomy. It rejects output measures and suggests a two-dimensional framework of using capacity and autonomy as a measure of executive branch quality. This framework explains the conundrum of why low-income countries are advised to reduce bureaucratic autonomy while high-income ones seek to increase it.

This commentary is the beginning of an effort to better measure governance, which at this point will amount to nothing more than an elaboration of the issue's complexity and the confused state of current discussions. Before we can measure good governance, however, we have to better conceptualize what it is.

The state, that is, the functioning of executive branches and their bureaucracies, has received relatively little attention in contemporary political science. Since the onset of the Third Wave of democratizations now more than a generation ago, the overwhelming emphasis in comparative politics has been on democracy, transitions to democracy, human rights, transitional justice, and the like. Studies of nondemocratic countries focus on issues like authoritarian persistence, meaning that the focus still remains the question of democracy in the long run or democratic transition. In other words, everyone is interested in studying political institutions that limit or check power—democratic accountability and rule of law—but very few people pay attention to the institution that accumulates and uses power, the state.

*Stanford University

The relative emphasis on checking institutions rather than power-deploying institutions is evident in the governance measures that have been developed in recent years. There are numerous measures of the quality of democracy like the Freedom House and Polity measures, as well as newer and very sophisticated ones like the Varieties of Democracy project led by Michael Coppedge, John Gerring, et al. We have fewer measures of Weberian bureaucracy—that is, the degree to which bureaucratic recruitment and promotion is merit based, functionally organized, based on technical qualifications, etc. One of the only studies to attempt to do this was by Peter Evans and James Rauch back in 2000, but their sample was limited to 30 odd countries and produced no time-series data. The Varieties of Democracy project is also collecting data on bureaucratic quality based on expert surveys. Other bureaucratic quality measures include the Bertelsmann Transformation Index, which "focuses on how effectively policymakers facilitate and steer development and transformation processes," and the proprietary Political Risk Service's Group International Country Risk Guide. Four of the six World Bank Institute's Worldwide Governance Indicators purport to measure aspects of state capacity (government effectiveness, regulatory quality, political stability and absence of violence, and control of corruption), but these are aggregates of other existing measures and it is not clear how they map onto the Weberian categories. For example, does a good absence of violence score mean that there is effective policing? I suspect that there isn't much by way of street crime or military coup attempts in North Korea. (These problems are also true of the Bank's internal CPIA scores.) Finally, Bo Rothstein's Quality of Governance Institute in Gothenberg has developed a set of measures of quality of governance for 136 countries worldwide, as well as a more detailed survey of 172 regions within the European Union. It is based again on expert surveys focusing on the degree of a state's impartiality, which Rothstein argues is a proxy for overall state quality.

The bias against thinking about state capacity is particularly strong among rational choice institutionalists. Most in this school begin with Mancur Olson's assumption that states are predatory and that the chief aim of political development is the creation of institutions like rule of law and accountability that limit the state's discretion. This school assumes that all states have the power to be predatory, and seldom raise the question of where state capacity comes from in the first place, or how it increases or decreases over time. Frankly, it would be very hard to develop a rational choice theory of state capacity, as capacity in any organization is so heavily influenced by norms, organizational culture, leadership, and other factors that do not easily fit into a model based on economic incentives.

In addition, there has been a large literature on public sector reform coming out of institutional economics, public administration, and from the communities of practice surrounding development agencies seeking to improve governance. The approaches favored by economists sought to conceptualize governance in a principal–agent framework, and have sought to control corruption and bad administration through manipulation of incentives. Many of the new approaches under this framework sought to bring market-like incentives into the public sector through creation of exit options, competition, manipulation of wage scales, shortening of accountability routes, and better methods of monitoring and accountability. Many of the techniques of New Public Management was in some sense an outgrowth of these approaches, though their applicability to developing world contexts has been questioned (Grindle 2004; Heady 1991; Pollitt and Bouckaert 2004; Schick 1998; World Bank 2004).

The existing measures of state quality or capacity have a number of limitations. There is an inherent weakness in expert surveys, especially when trying to create time-series data. As the concept of good governance is not well established, different experts may intend different things when responding to the same survey question. For example, there is an important difference between clientelism and outright corruption; in the former there is true reciprocity between patron and client, whereas in the latter there is no obligation on the part of the corrupt official to give anything back. The economic impact of corruption varies tremendously depending on whether the corruption "tax" is 10% or 50%, and the quality and nature of the services that clients get in return. In China, for example, corruption seems to be pervasive, but the tax rate is lower and the service provision rate much higher than in, say, sub-Saharan Africa. None of the existing corruption surveys are, as far as I know, able to make distinctions of this sort.

Bo Rothstein makes a number of persuasive arguments that impartiality ought to be the core measure of the quality of government. However, it would seem entirely possible that a state could be highly impartial and still lack the capacity and/or autonomy to effectively deliver services. Rothstein argues that impartiality implies the existence of sufficient capacity. This may be true, but it is something that needs to be empirically verified rather than simply asserted.

In addition, there are a number of Rule of Law measures that relate to bureaucratic quality, such as those published by the ABA Rule of Law Initiative and the World Justice Project's Rule of Law Index. Some

Chinese scholars have tried to measure the spread of rule-based decision making by measuring the number of administrative cases filed against government agencies, as well as the percentage of such cases that are won by the plaintiffs.

The rule of law is defined differently by different scholars and can mean, alternatively, law and order, property rights and contract enforcement, observance of substantive Western norms of human rights, and constitutional constraints on the power of the executive (Kleinfeld 2006). Some scholars have distinguished between rule by law, in which the executive uses law and bureaucracy as an instrument of power, and rule of law, in which the executive is itself constrained by the same laws that apply to everyone else. In many respects rule by law overlaps with state quality, as we want states to operate by general, transparent, impartial, and predictable rules. Rule of law in the narrow sense of constitutional constraints on the executive, on the other hand, is closely associated with democracy. The Prussian/German Rechtsstaat in the 19th century, Meiji Japan, and contemporary China were all authoritarian states that could be said to have rule by law but not rule of law. This means that certain aspects of rule of law would be useful as measures of state quality. On the other hand, many rule of law measures measure what is measurable rather than the underlying quality of law, so we would need to be careful in selecting them.

Definitions

As a starting point, I am going to define governance as a government's ability to make and enforce rules, and to deliver services, regardless of whether that government is democratic or not. I am more interested in what Michael Mann labels "infrastructural" rather than "despotic" power (Mann 1984). The reason I am excluding democratic accountability from the definition of governance is that we will later want to be able to theorize the relationship between governance and democracy. The current orthodoxy in the development community is that democracy and good governance are mutually supportive. I would argue that this is more of a theory than an empirically demonstrated fact, and that we cannot empirically demonstrate the connection if we define one to include the other.

In this initial conceptualization, the quality of governance is different from the ends that governance is meant to fulfill. That is, governance is about the performance of agents in carrying out the wishes of principals, and not about the goals that principals set. The government is an organization that can do its functions better or

worse; governance is thus about execution, or what has traditionally fallen within the domain of public administration, as opposed to politics or public policy.[1] An authoritarian regime can be well governed, just as a democracy can be maladministered. (As we will see below, this distinction cannot always be maintained quite so neatly; principals can set self-undermining tasks for their agents.)

As Rothstein (2011) points out, it is not so easy to separate governance as implementation from the normative ends that government is meant to serve. It is not clear that a well-governed state is one that has ruthlessly efficient concentration camp guards as opposed to bribable ones. On the other hand, once one starts to introduce substantive ends as criteria for good government, it is hard to know where and when to stop. As Rothstein points out, the existing Worldwide Governance Indicators embed a number of normative policy preferences (e.g., less rather than more regulation) that color the final results. Would we want to argue that the U.S. military is a low-quality one because it does things we disapprove of, say, invading Iraq?

Rothstein argues that use of the criterion of impartiality solves this problem as it is both normative and embeds what most people understand by "good government." However, for reasons I will elaborate below, I don't think that impartiality by itself is a sufficient metric.[2] I want to put the normative question to the side for the time being, however, particularly because I am interested in developing measures that will work for both authoritarian and democratic regimes. Focusing on an extreme case like concentration camps should not distract us from the fact that there are many valence issues like provision of education, health, or public safety that are shared by virtually all governments, in which a more instrumental view of quality of governance will suffice.

If we accept this definition of the object we are trying to study, then there are at least four broad approaches to evaluating the quality of governance: procedural measures, input measures, output measures, and measures of bureaucratic autonomy.

Procedural Measures

The most classic effort to define governance in terms of procedures was Max Weber's famous characterization of modern bureaucracy in *Economy and Society* (Weber 1978, 220–221). We continue to use the term "Weberian bureaucracy" as an ideal type to which we hope highly corrupt, neo-patrimonial states will eventually conform. It might be useful to review Weber's conditions here:

1. Bureaucrats are personally free and subject to authority only within a defined area.
2. They are organized into a clearly defined hierarchy of offices.
3. Each office has a defined sphere of competence.
4. Offices are filled by free contractual relationship.
5. Candidates are selected on basis of technical qualifications.
6. Bureaucrats are remunerated by fixed salaries.
7. The office is treated as the sole occupation of the incumbent.
8. The office constitutes a career.
9. There is a separation between ownership and management.
10. Officials are subject to strict discipline and control.

Conditions 1–5 and 9 are probably at the core of what people think of when they talk about "modern bureaucracy": They clearly delineate such an organization from the kinds of venal or patrimonial office that existed in Europe under the Old Regime, or that exist in contemporary neo-patrimonial developing countries today. However, characteristics 6, 7, 8, and 10 are more problematic. Condition 6, fixed salaries, is not compatible with the kinds of incentives often offered bureaucrats under New Public Management. Conditions 7 and 8 are not true of many mid-level officials in contemporary America, in both the public and private sectors. One could say that the United States fails to live up to the Weberian ideal, but it does not seem likely that the quality of bureaucracy in the United States would improve if it were impossible for talented individuals from the private sector or the academy to serve in government for periods of time. And condition 10 is incompatible with civil service protection, which during the Progressive Era was seen as a hallmark of the modern bureaucracy that was replacing the patronage system. More importantly, condition 10 suggests that bureaucrats are simply robotic agents whose only purpose is to do the bidding of principals. The idea of bureaucratic autonomy—the notion that bureaucrats themselves can shape goals and define tasks independently of the wishes of the principals—is not possible under the Weberian definition.

Nonetheless, certain procedural measures would remain at the core of any measure of quality of governance. One would want to know whether bureaucrats are recruited and promoted on the basis of merit or political patronage, what level of technical expertise they are required to possess, and the overall level of formality in bureaucratic procedure.

Capacity Measures

The problem with all procedural definitions of bureaucracy is that the procedures, however defined, may not actually correlate with the positive outcomes expected from governments. We assume that a Weberian bureaucracy will produce better services than one that is highly discretionary and patrimonial, yet there may be circumstances where the latter's lack of rules result in faster and better tailored responses. Enforcement power is not part of Weber's definition; it is possible to have an impersonal, merit-based bureaucracy that nonetheless is extremely poor at getting things done. To say that a bureaucrat is selected on the basis of "merit" does not define merit, nor does it explain whether the official's skills will be renewed in light of changing conditions or technology.

The most commonly used measure of capacity is extractive capacity, measured in terms of tax extraction. Tax extraction measures capacity in two ways: First, it takes capacity, however generated, in order to extract taxes; second, successful tax extraction provides resources that enable the government to operate in other domains. Tax extraction rates can be measured both by the percentage of taxes to gross domestic product, as well as by the nature of taxation—that is, whether it is based on income or wealth, or indirect taxation (as income and wealth taxes are much more difficult to extract than indirect taxes).

While tax extraction is a reasonable starting point for measuring capacity, it has several important limitations:

1. There is a difference between extractive potential and actual extraction rates. Actual tax rates are set not just by extractive potential, but by policy choices regarding the optimal rate and types of taxation.[3] The United States proved it had the potential to extract significantly higher levels of taxes during the two World Wars, because it had an overriding national interest in doing so. The peacetime level reflects normative preferences for the optimal size of government, which may vary between countries of identical potential capacity.[4]

2. A given level of taxation does not necessarily translate into the efficient use of tax revenues. Revenues can be wasted on poor administration, unproductive transfers, or outright corruption. Bureaucratic outputs are the result not just of resource inputs, but of things like organizational culture. Judith Tendler has written about a poor and underresourced state in Northeast Brazil that nonetheless achieved very good governance outcomes (Tendler 1997).[5]

3. For many countries, government revenues are based on resource rents or international transfers rather than domestic taxation. In many countries such rents and transfers constitute the vast majority

of government revenues. One could argue that if taxation is going to
be used as a measure of state capacity, then resource rents ought to
be excluded.

Tax extraction rates are hardly the only possible measures of state
capacity. States perform a whole variety of functions, any one of which
can be used as a proxy for state capacity as a whole. Taxation is a useful
proxy for general capacity because it is a necessary function of all
states, and one for which considerable data exist. In an ongoing doc-
toral research, Melissa Lee and Nan Zhang have suggested using the
capacity to generate accurate census data as an alternative proxy for
capacity, as population registration is a very basic state function.

Beyond taxation, another critical measure of capacity is the level of
education and professionalization of government officials. Central
banks in the early 21st century across the developing world are incom-
parably better run than they were in the lead-up to the debt crises of
the 1980s in Latin America and sub-Saharan Africa, due in part to the
significantly higher degree of professionalism in their staffing. A key
aspect of state building in the United States during the Progressive Era
was the replacement of incompetent political patronage appointees
with university-trained agronomists, engineers, and economists.

A focus on the degree of professionalization of the bureaucracy par-
tially solves the problem of how to measure levels of corruption, a
measure that is not dependent on expert or perception surveys. All
professional education (with the possible exception of business
schools) embeds a strong normative element in which service to one's
profession and broader public goals is paramount. A doctor, for
example, is supposed to act primarily in the interests of the patient
rather than seeking first to maximize his or her individual benefit. Of
course, all professionals are also selfish individuals who can act in a
corrupt manner. But in modern organizations we trust highly edu-
cated professionals with a much higher degree of discretion because
we assume or hope that they will be guided by internal norms in cases
where their behavior cannot be monitored from the outside.

As state capacity varies substantially across functions, levels of govern-
ment, and regions, one would ideally want capacity measures for *all*
major government agencies. In Brazil, for example, it has been widely
recognized that certain "islands of excellence" exist within the Brazil-
ian state that would be missed by an aggregate measure. Thus, an
article by Katherine Bersch, Sergio Praça, and Matthew Taylor devel-
ops capacity measures across more than 300 different Brazilian federal
agencies (Bersch, Praça, and Taylor 2012). Obviously, this kind of data
does not exist for many countries, and even in Brazil the authors do

not have similar statistics for capacity at the state, local, and municipal levels where a great deal of governance happens. For evaluating a country like China, it would be very important to generate this type of disaggregated data, as there is a widely held perception that the quality of governance varies enormously across the different levels and functions of government.

As a kind of compromise between an unachievable ideal of fully disaggregated capacity data and a limited aggregate measure, it might be possible to specify a subset of government functions on which data should be collected. This could be a set of functions theoretically performed by all governments (e.g., macroeconomic policy management, basic law and order, primary and secondary education, population registration), or it could incorporate data on how expansive the functions performed are (e.g., giving extra credit if a government is able to, say, regulate pharmaceuticals).

Output Measures

Good procedures and strong capacity are not ends in themselves. We want governments to do things like provide schooling and public health, public security, and national defense. This suggests an alternative measure of government quality, a measure of final output. One could look at literacy, primary and secondary education test scores, or various measures of health to get some idea as to how governments are performing.

Attractive as output measures sound, there are several big and, in my view, decisive drawbacks to their use. First and most important, outputs like health or education are not simply the consequences of public action; the public sector interacts with the environment around it and the society it is dealing with to produce results. For example, the Coleman Report on U.S. education in the 1960s showed that educational outcomes depended much more strongly on factors like friends and family of students than they did on public sector inputs to education (Coleman 1966). Joel Migdal's model of weak states and strong societies suggests that a government's ability to penetrate or regulate a society depends on the ratio of two factors, state capacity and the self-organization of the underlying society. Two states could have equal regulatory capacity but unequal regulatory outcomes because the society in one is far better organized to resist state penetration than the other (Migdal 1988).

A second problem is that measuring output is itself problematic methodologically. Public sectors produce primarily services, which can be

notoriously hard to measure. For example, standardized test scores, a common way of evaluating educational outcomes, have long been under attack as poor measures of education, and for creating incentives to "teach to the test." Measures of rule of law, like time to trial, rate of case clearances, etc., say nothing about the quality of the justice being produced by a legal system.

Finally, outcome measures cannot be so easily divorced from procedural and normative measures. A police state may succeed in controlling street crime by massively arresting and torturing suspects, yet most believers in liberal democracy would accept a higher degree of crime in exchange for procedural protections of individual rights. Even if one is morally neutral about whether torture is justified as a police method, one would want to know whether it is employed routinely in evaluating a government.

My own sense is that the problem of the tainting of output measures by exogenous factors means that they should not be used as state quality measures in the first place. One could employ a variety of econometric techniques to control for these exogenous factors, but that entails another layer of complexity and problems. In fact, it might be better to leave output as a dependent variable to be explained by state quality, rather than being a measure of capacity in itself. If output is not a valid measure of state quality, it implies that we also cannot generate useful measures of government efficiency as a measure of state quality, as the latter represents a ratio of state inputs to outcomes.

Bureaucratic Autonomy

A final measure of the quality of government is the degree of bureaucratic autonomy possessed by the different components of the state. Samuel Huntington makes autonomy one of his four criteria of institutionalization; highly institutionalized political systems have bureaucracies with high autonomy. The opposite of autonomy in Huntington's terminology is subordination (Huntington 2006).[6]

Autonomy, properly speaking, refers to the manner in which the political principal issues mandates to the bureaucrats who act as its agent. No bureaucracy has the authority to define its own mandates, regardless of whether the regime is democratic or authoritarian. But there are a wide variety of ways in which mandates can be issued. Ideally, the principal should set a broad mandate to the agent, for example, procurement of an advanced strike fighter. But the principal can also issue many other mandates as well regarding the way in which to carry out the broad mandate, such as purchasing a strike fighter using contrac-

tors that increase employment in Congressional districts X and Y, or through minority and women-owned businesses, or to achieve Z degree of performance desired by a rival service. In other cases the principal can issue mandates regarding the bureaucracy's recruitment and promotion of personnel, requiring that they hire certain individuals, or else setting detailed rules for personnel management.

Political principals often issue frequently overlapping and sometimes downright contradictory mandates. Indeed, there can be multiple principals in many political systems, that is, political authorities with equal legitimacy able to issue potentially contradictory mandates. State-owned utilities, for example, often have mandates to simultaneously do cost recovery, universal service to the poor, and efficient pricing to business clients, each promoted by a different part of the political system. These different mandates obviously cannot be simultaneously achieved, and generate bureaucratic dysfunctionality. Amtrak could become a profitable and efficient railway if it were not under Congressional mandates to serve various low-volume rural communities. In China there are often duplicate functional agencies, one reporting through a chain of command that goes through national ministries, the other reporting to municipal or provincial governments; it is not always clear how conflicts between them are to be resolved.

Autonomy therefore is inversely related to the number and nature of the mandates issued by the principal. The fewer and more general the mandates, the greater autonomy the bureaucracy possesses. A completely autonomous bureaucracy gets no mandates at all but sets its own goals independently of the political principal. Conversely, a non-autonomous or subordinated bureaucracy is micromanaged by the principal, which establishes detailed rules that the agent must follow.

An appropriate degree of bureaucratic autonomy does not mean that bureaucrats should be isolated from their societies or make decisions at odds with citizen demands. Indeed, if the general mandate is to provide high-quality services in health or education, the bureaucracy would need considerable feedback and criticism from the citizens that it is trying to serve. It also does not exclude extensive collaboration with private sector or civil society organizations in service delivery. Indeed, an appropriately autonomous bureaucracy should be able to make judgment calls as to when and where to engage in such collaborations.

It would seem that the relationship between autonomy and quality of government would look like an inverted U (see Figure 1). At one extreme, that of complete subordination, the bureaucracy has no room for discretion or independent judgment, and is completely bound by

FIGURE 1
Bureaucratic Autonomy and Quality of Government
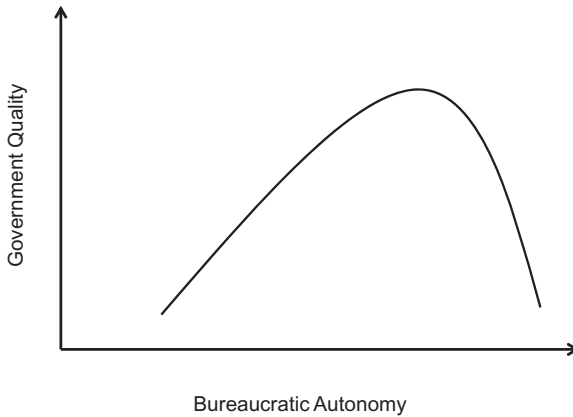
detailed rules set by the political principal. At the other end of the x-axis, that of complete autonomy, governance outcomes would also be very bad, because the bureaucracy has escaped all political control and sets not just internal procedures but its goals as well. This is basically the idea contained in Peter Evans' concept of "embedded autonomy": Bureaucrats need to be shielded from certain influences of social actors, but also subordinate to the society with regard to larger goals (Evans 1995).

There are myriad examples of excessive subordination leading to poor performance. One of the worst forms is when bureaucracies lose control over internal recruitment and promotion to the political authorities and are staffed entirely by political appointees. This is in effect what happens in clientelistic political systems. But even in the absence of clientelism, bureaucracies can be excessively slow moving and indecisive because they are excessively rule bound. However, the curve in Figure 1 slopes downward at the left end of the x-axis, which represents full autonomy. Both Imperial Germany and Japan in the periods before World War I and World War II, respectively, suffered from this problem. Both countries had developed very high quality, autonomous bureaucracies, particularly their military services, which then took over from the political authorities the task of formulating foreign policy.

The inflection point of the curve in Figure 1 is shifted to the right, however, due to a general recognition that the dangers of excessive micromanagement are greater than those posed by excessive

autonomy. A high degree of autonomy is what permits innovation, experimentation, and risk taking in a bureaucracy. In Daniel Carpenter's *The Forging of Bureaucratic Autonomy*, both the Post Office and the U.S. Forest Service are portrayed as high-quality autonomous bureaucracies during the Progressive Era precisely because they innovated and devised agendas not strictly spelled out by Congress (Carpenter 2001). This same insight is embedded in the evolution of the U.S. Army's basic field manual for combined arms operations, FM 100-5. In rethinking combined arms doctrine in light of the Vietnam War, the drafters of the manual shifted emphasis from centralized command and control to more flexible Mission Orders under which the commander only set broad goals, and devolved implementation to the lowest possible echelon of the command structure. The latter were in other words agents who were permitted a high degree of autonomy, which included toleration of failure if they sought to innovate or experiment.[7] More broadly, one could argue that modern private sector organizations have evolved over time away from rigid "Taylorite" hierarchies reflecting strict Weberian criteria to more flexible, flatter organizations that delegate far more authority to lower levels of the organization.

If an appropriate degree of bureaucratic autonomy is an important characteristic of high-quality government, then neither the Weberian nor the principal–agent models can stand intact as frameworks for understanding how bureaucracies ought to work. The Weberian model, as noted earlier, assumes that bureaucrats are essentially rule-bound implementers of decisions made by political authorities; they may have technical capacity but they do not have the authority to set agendas independently. The principal–agent framework is inadequate as well because it, too, assumes that agents are simply tools of the principals, whereas in a good bureaucracy authority often flows in the reverse direction, from the agent to the principal (the latter point, basic in an older tradition of public administration, is made by Herbert Simon; Simon 1957).

How do we measure bureaucratic autonomy? I believe that this is one of the most central and difficult issues in constructing a good measure of government quality. The most common approach is to use expert surveys in which experts are asked to evaluate the autonomy of a given bureaucracy. Expert surveys are particularly problematic in this area because the very concept of autonomy has been poorly specified, and it is not clear exactly what it is that experts are being asked to judge. Do they have adequate criteria for judging what proper and improper mandates are, or to look for multiple or conflicting mandates as a measure of subordination?

It would be nice therefore to have a more objective measure of autonomy. As autonomy is the opposite of subordination, one could use the degree of clientelism or political interference in bureaucratic operations as a measure. One could look, for example, at the relative number of classified versus political positions in a bureaucracy. However, this gets at only one type of subordination related to personnel. Political principals can hamstring bureaucracies by issuing multiple contradictory mandates that have nothing to do with staffing, or by setting excessively detailed rules for bureaucratic behavior.
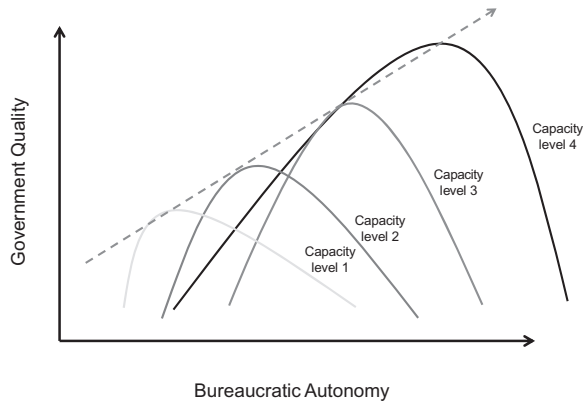
Capacity *and* Autonomy

It would seem to be the case that the quality of government is the result of an interaction between capacity *and* autonomy. That is, more or less autonomy can be a good or bad thing depending on how much underlying capacity a bureaucracy has. If an agency were full of incompetent, self-dealing political appointees, one would want to limit their discretion and subject them to clear rules. The assertion embedded in Figure 1 that the optimal amount of autonomy is shifted to the right is true only in high-capacity countries. In very low capacity countries, the opposite would be the case: One would want to circumscribe the behavior of government officials with more rather than fewer rules because one could not trust them to exercise good judgment or refrain from corrupt behavior. This is why Robert Klitgaard coined the formula Corruption = Discretion – Accountability (Klitgaard 1988). This is also why development agencies have been advising poor countries to limit bureaucratic discretion in recent years. On the other hand, if the same agency were full of professionals with graduate degrees from internationally recognized schools, one would not just feel safer granting them considerable autonomy, but would actually want to *reduce* rule boundedness in hopes of encouraging innovative behavior.

Figure 2 illustrates how the optimal autonomy curves would differ for four hypothetical countries of differing levels of capacity. For each, the curve slopes downward at the extremes, because every bureaucracy can have too much or too little autonomy. But the lower-capacity countries have their inflection points shifted to the left, while they are shifted right for higher-capacity countries.

One can control the behavior of an agent either through explicit formal rules and incentives or through informal norms and habits. Of the two, the latter involves substantially lower transaction costs. Many professionals are basically self-regulated, due to the fact that (1) it is hard for people outside their profession to judge the quality of their work and (2) part of their education, as noted previously, consists of socialization

FIGURE 2
Optimal Levels of Autonomy for Differing Levels of Capacity

to certain professional norms that seek to preclude certain types of self-seeking behavior. The higher the capacity of a bureaucracy, then, the more autonomy one would want to grant them. In judging the quality of government, therefore, we would want to know about both the capacity *and* the autonomy of the bureaucrats.

One would want, in other words, to be able to empirically locate the agency on the matrix shown in Figure 3. The line sloping downward and to the left represents the line drawn through the inflection points of Figure 2, representing optimal levels of autonomy for a given level of capacity. Bureaucracies that were to the left of the line would be hobbled by excessive rules; those to the right of it with excessive discretion. For the past decade, international donors have been advising developing countries to decrease the amount of discretion in the behavior of their bureaucracies. From Figure 3 it would appear that this is only contingently good advice; in a high-capacity state, one would like to have more rather than less discretion.

The framework in Figure 3 suggests that there are two quite separate approaches to public sector reform. One always wants to move up the y-axis to higher capacity, particularly with regard to the professionalism of the public service. This, however, is not something that can be done easily, and it is not something that can in any case be accomplished in a short period of time. If a country cannot significantly upgrade capacity in the short run, one would want to shift the degree of autonomy toward the sloping line. This would mean moving toward the left in a low-capacity country, and toward the right in a
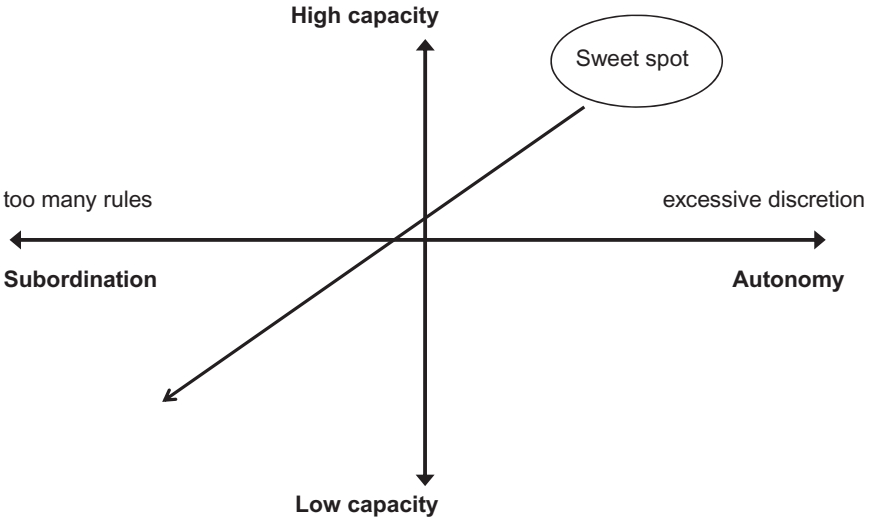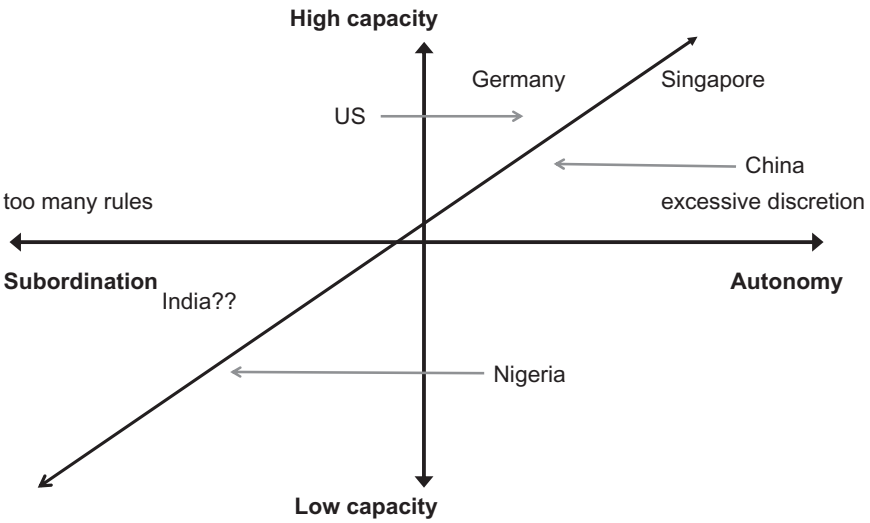
FIGURE 3
Autonomy and Capacity

High capacity

Sweet spot

too many rules   excessive discretion

Subordination   **Autonomy**

Low capacity

FIGURE 4
Reform Paths

High capacity

Germany   Singapore

US

China

too many rules   excessive discretion

Subordination   **Autonomy**
India??

Nigeria

Low capacity

high-capacity country. Figure 4 contains some hypothesized positions, aggregated across the whole government, for different countries, and suggests that while Nigeria and China need to move left, the United States needs to move right. However, China needs to end up at a point with significantly more autonomy than Nigeria because of its much higher capacity.

Trying to locate India on this matrix demonstrates some of the complexities of this analysis. India is famous both for high levels of corruption and clientelism, and for simultaneously having excessive rules and bureaucratic red tape. India clearly needs much greater state capacity across the board. But does it need more or less autonomy? The answer to the latter question is probably both, dependent on specific context. Given the recent scandal, the agency handling spectrum auctions needs to be subjected to much stricter rules; on the other hand, the Hyderabad Municipal Water Authority needs to be relieved of its multiple and conflicting political mandates if it is to function properly. This then suggests why devising single aggregated measures for quality of governance can be inadequate and misleading.

Conclusion

It is clear that in evaluating the quality of governance in large, complex countries like China or the United States, the existing quantitative measures are woefully inadequate. If we are to establish desiderata for better ones, we would have to answer the following questions:

- If we are to use procedural measures of government quality, which on Weber's list do we want to keep?
- For how many countries could we collect disaggregated capacity data?
- If we cannot collect a full set of capacity measures, what are the best proxies for aggregate capacity? Beyond tax extraction levels, can we come up with measures of bureaucratic professionalization?
- How do we distinguish between actual and potential capacity, with regard to a commonly used measure like tax extraction?
- How, exactly, are we going to define bureaucratic autonomy, and what measures are available as proxies for it?
- How important is it to have quantitative measures at all, as opposed to qualitative descriptions of process, or else case studies of particular areas of governance?

This commentary does not pretend to answer these questions, but only to serve as a basis for discussion. As we cannot measure what we

cannot adequately conceptualize, we have to start with the concept first. I have laid out two separate dimensions of governance, capacity and autonomy, and suggested some of the components that make them up. Capacity, in particular, consists of both resources and the degree of professionalization of bureaucratic staff. I have further posited that quality of governance is ultimately a function of the interaction of capacity and autonomy, and that either one independently will be inadequate as a measure of government quality. Finally, I have also suggested that states need to be disaggregated into their component parts, by function, region, and level of government, and that we need both capacity and autonomy measures for all of these components. Obviously, this volume of data does not exist for most countries, and may not exist for any country. How much could be generated? It might be useful to start with a large, relatively data-rich country like the United States, and see how far we could get.

Acknowledgments

Notes

1. This distinction was made in Woodrow Wilson's famous article (Wilson 1887). It is also made in Max Weber's equally famous essay, "Politics as a Vocation" (Weber 1946).
2. One would have to say that a concentration camp guard who executed everyone he was ordered to kill was more impartial than one who played favorites or spared certain individuals in return for bribes or sexual favors. This points to the difference between impartiality of policies compared to the impartiality of the way in which policies are executed.
3. Mancur Olson and others in the rational choice tradition argue that states are predatory and that all states will seek to tax at a maximal rate, subject only to limitations on capacity, and the time discount rates of the sovereign. There is, however, considerable historical evidence that this is not true, and that states have deliberately taxed well below their theoretical capacities for a variety of reasons (Fukuyama 2011, 303–305; Olson 1993).
4. Marcus Kurtz in his forthcoming book on the state in Latin America makes use of this distinction (Kurtz 2013).
5. On the general importance of organizational culture, see DiIulio (1994) and Wilson (1989).
6. One could use the term "accountability" as the antonym for autonomy; however, accountability has certain normative implications that subordination does not.

7. This field manual was based on the operational doctrine that had been developed by the German army from the end of World War I through the beginning of World War II; Mission Orders are an American version of *Aufstragstaktik* (Fukuyama and Shulsky 1997).

References

Bersch, Katherine, Sergio Praça, and Matthey Taylor. 2012. "An Archipelago of Excellence? Autonomous State Capacity among Brazilian Federal Agencies." Unpublished paper.

Carpenter, Daniel. 2001. *The Forging of Bureaucratic Autonomy: Reputations, Networks, and Policy Innovation in Executive Agencies, 1862–1928*. Princeton, NJ: Princeton University Press.

Coleman, James S. 1966. *Equality of Educational Opportunity*. Washington, DC: U.S. Department of Health, Education and Welfare.

DiIulio, John J. 1994. "Principled Agents: The Cultural Bases of Behavior in a Federal Government Bureaucracy." *Journal of Public Administration Reseach and Theory* 4 (3): 277–320.

Evans, Peter. 1995. *Embedded Autonomy*. Princeton, NJ: Princeton University Press.

Fukuyama, Francis. 2011. *The Origins of Political Order: From Prehuman Times to the French Revolution*. New York: Farrar, Straus and Giroux.

Fukuyama, Francis, and Abram Shulsky. 1997. *The Virtual Corporation and Army Organization*. Santa Monica, CA: Rand Corporation.

Grindle, Merilee S. 2004. "Good Enough Governance: Poverty Reduction and Reform in Developing Countries." *Governance* 17 (4): 525–548.

Heady, Ferrel. 1991. *Public Administration: A Comparative Perspective*. New York: Marcel Dekker.

Huntington, Samuel P. 2006. *Political Order in Changing Societies*. New Haven, CT: Yale University Press.

Kaufmann, Daniel, and Aart Kraay. 2009. *Governance Matters VIII: Aggregate and Individual Governance Indicators, 1996–2008*. Washington, DC: World Bank Institute.

Kleinfeld, Rachel. 2006. "Competing Definitions of the Rule of Law." In *Promoting the Rule of Law Abroad: In Search of Knowledge*, ed. T. Carothers. Washington, DC: Carnegie Endowment, 31–74.

Klitgaard, Robert. 1988. *Controlling Corruption*. Berkeley, CA: University of California.

Kurtz, Marcus. 2013. *Latin American State Building in Comparative Perspective*. New York: Cambridge University Press.

Mann, Michael. 1984. "The Autonomous Power of the State: Its Origins, Mechanisms, and Results." *European Journal of Sociology* 25 (2): 185–213.

Migdal, Joel S. 1988. *Strong Societies and Weak States: State-Society Relations and State Capabilities in the Third World*. Princeton, NJ: Princeton University Press.

Olson, Mancur. 1993. "Dictatorship, Democracy, and Development." *American Political Science Review* 87 (9): 567–576.

Pollitt, Christopher, and Geert Bouckaert. 2004. *Public Management Reform: A Comparative Analysis*. 2nd ed. New York: Oxford University Press.

Rothstein, Bo. 2011. *The Quality of Government: Corruption, Social Trust, and Inequality in International Perspective*. Chicago: University of Chicago Press.

Schick, Allen. 1998. "Why Most Developing Countries Should Not Try New Zealand Reforms." *World Bank Research Observer* 13 (8): 123–131.

Simon, Herbert. 1957. *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*. New York: Free Press.

Tendler, Judith. 1997. *Good Government in the Tropics*. Baltimore, MD: Johns Hopkins University Press.

Weber, Max. 1946. *From Max Weber: Essays in Sociology*. New York: Oxford University Press.

Weber, Max. 1978. *Economy and Society*. Berkeley, CA: University of California Press.

Wilson, James Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books.

Wilson, Woodrow. 1887. "The Study of Administration." *Political Science Quarterly* 2 (2): 197–222.

World Bank. 2004. *World Development Report 2004: Making Services Work for Poor People*. Washington, DC: World Bank.

Appendix: The Inadequacy of Existing Measures of China's Quality of Government

If we accept the fact that quality of government is a mixture capacity and autonomy, and that governments are themselves complex collections of organizations, then it becomes clear that existing measures of governance are highly inadequate. The Worldwide Governance Indicators produced by the World Bank Institute (Kaufmann and Kraay 2009), as well as finer-grained measures like Transparency International's Corruption Perception Index, treat single sovereign nations as the unit of analysis. Yet it is obvious that the quality of governance varies enormously within countries, both by specific government function and by region. Moreover, one cannot look at governance problems at one level only; many occur because of interactions between levels of governments. A poor national government can reduce the performance of a good local one, and vice versa.

The problem of single country indicators is evident when we consider something like Transparency International's Corruption Perception Index. The 2011 index lists China as the 75th most corrupt country in the world. It does a bit better than Brazil and Tunisia (both #73), it tied with Romania, and it is just slightly better than Gambia and El Salvador. Yet this number is virtually meaningless because it does not take account of the diversity of outcomes within China. It is widely believed in China, for example, that local governments there are much more corrupt than higher-level ones. We do not in fact know whether this is true or not. Corruption varies not just by level of government, but by region and by function; the railroad ministry is very different from, say, the Central Bank.

There is also something very strange about the Worldwide Governance Indicators rankings of China (see Table A1).

China's low rankings for Voice and Accountability and Rule of Law are not surprising, given that no one argues either of these are China's strong suit. The other four measures relate to what we are defining as governance. While both the score and ranking for government effectiveness are higher

TABLE A1
China's 2010 Performance, Worldwide Governance Indicators (Scores Range from –2.5 [Weak] to 2.5 [Strong])

Category	Score	Percentile
Voice and accountability	−1.6	5
Political stability/no violence	−0.77	24
Government effectiveness	0.12	60
Regulatory quality	−0.23	45
Rule of law	−0.35	44
Control of corruption	−0.60	33

than for any other measure, China still places only in the 60th percentile. But what possible meaning can such a figure have? Clearly many local Chinese government authorities have huge problems; on the other hand, others perform far better. In my purely subjective estimation, the effectiveness of China's national government with regard to macroeconomic management of a hugely complex modernization process over the past three decades has been nothing short of miraculous, given the fact that China was not just managing an existing set of institutions, but also transforming them in a more market-friendly direction. Its performance since the Asian financial crisis has arguably been better than that of the United States, which nonetheless ranks in the 90th percentile.

In terms of the three categories above, what do existing measures of governance measure? In the case of the WBI Worldwide Governance Measures (WWGM), it is hard to say, because they are an aggregate of many other measures. Many of them are perception surveys or expert estimates, which often reflect output measures, but may also include evaluations of procedures and capacity. It is not clear whether any of the WWGM components explicitly seek to measure bureaucratic autonomy. Presumably categories like Political Stability/Control of Violence are exclusively output measures (where China's low 24th percentile ranking seems a bit bizarre). The Rule of Law measure has big problems, beginning with the lack of definition of what is being measured. If rule of law is defined as constraints on the executive, China should rank even lower than it does as there are no real legal constraints on the behavior of the Chinese Communist Party. If on the other hand this category means something more like rule by law (which would make it a component of governance), the ranking should be considerably higher. Most Rule of Law measures tend to be related to procedures or capacity rather than output, because the output of a legal system is so hard to measure. But we actually have no idea what the Chinese numbers actually mean or purport to measure.

Francis Fukuyama is the Olivier Nomellini Senior Fellow at the Freeman Spogli Institute for International Studies at Stanford University. Dr. Fukuyama has written widely on issues relating to democratization and international political economy. His book, *The End of History and the Last Man,* was published by Free Press in 1992. His most recent book, *The Origins of Political Order*, was published by Farrar, Straus, and Giroux in 2011. Other books include *America at the Crossroads: Democracy, Power, and the Neoconservative Legacy*, and *Falling Behind: Explaining the Development Gap between Latin America and the United States.*