Beyond State Capacity: Bureaucratic Performance, Policy Implementation, and Reform

Martin J. Williams *

This version: 5 October, 2017

Abstract

State capacity has become a default concept across the social sciences for studying government bureaucracies and how to improve them. I argue that “capacity” is a coarse and unhelpful way to understand bureaucratic performance and policy implementation, because it obscures both the mechanisms and the contingencies of bureaucratic behavior. Whereas capacity refers to bureaucracies’ hypothetical potential rather than their actual actions, a vast literature emphasizes that bureaucracies perform at less than their hypothetical potential due to internal information and incentive problems created by their collective nature, and the constraints and uncertainty imposed by their multiple political principals. Moreover, when discussing the actual implementation of specific policies, general notions of state capacity quickly collapse into a more complex set of explanations rooted in specific political, bureaucratic, and contextual contingencies. While the term may be a convenient shorthand to reduce the complexities of bureaucratic performance, policy implementation, and long-term state formation into one term, it is rarely a useful way to understand how bureaucracies function or how to reform them. Scholars should focus on retrospective bureaucratic performance rather than prospective potential, and ground analysis of policy implementation in specific contexts.

*Associate Professor in Public Management, University of Oxford, Blavatnik School of Government. Email: martin.williams@bsg.ox.ac.uk. I am grateful for comments from Yuen Yuen Ang, Michael Bernhard, Yanilda María Gonzalez, Tobias Haque, Dan Honig, David Jacobstein, Julien Labonne, Brian Levy, Peace Medic, Dan Rogger, Bo Rothstein, and seminar participants at Johns Hopkins University. Any remaining mistakes are my own.
Beyond State Capacity:
Bureaucratic Performance, Policy Implementation, and Reform

People (i.e., individuals) have goals; collectivities of people do not.
- Cyert and March (1963, 30)

Congress Is a “They”, Not an “It”
- Shepsle (1992, 239)

1 Introduction

Policy implementation and bureaucratic performance have become central issues for studies of service delivery, governance, and politics, particularly in developing countries. These discussions frequently revolve around the concept of “state capacity” - the ability of the state bureaucracy to implement government’s policy choices. In recent years there has been an explosion of literature on the conceptualization, determinants, measurement, and consequences of state capacity (Englehart 2009; Besley and Persson 2011; Lee et al 2014; Harbers 2015; Muralidharan et al 2016; Bersch et al forthcoming). While the idea originated in the macro-historical literature on state formation (Mann 1984), state capacity has become a catch-all term used to discuss issues of policy implementation, bureaucratic performance, and reform.

However, “capacity” is a misleading metaphor for the functioning of government bureaucracies, because it misrepresents the mechanisms of bureaucratic performance and policy implementation, and obscures the contingency of performance and implementation on the specifics of politics, policies, and contexts. The term misrepresents the mechanisms of bureaucratic performance because it conceives of bureaucratic action in terms of a bureaucracy’s hypothetical ability to implement policies. While the notional potential of a bureaucracy to implement policy may be equivalent to its actual performance in the simplest case of a unitary agent implementing well-defined policy choices to the best of its ability, the dominant feature of actually existing bureaucracies is that they are composed of and directed by a multiplicity of actors. Organizations are collectivities composed of many agents with different preferences and incentives, and their efficient operation depends largely on resolving the resulting problems of information and incentives (Garicano
and credibility and clarity (Gibbons and Henderson 2013). Similarly, government policy decisions are not the unambiguous command of a single political principal, but are unstable and incomplete expressions of constantly shifting collective choices among multiple political principals (Wilson 1989; Shepsle 1992).

While individuals may have specific capacities, conceiving of organizations as having capacities thus obscures perhaps the most salient characteristic of organizations: that they are collectives. As Shepsle (2002, 339) writes in his classic polemic against the concept of legislative intent, another widely used term that is not grounded in a rigorous understanding of collective behavior: “To claim otherwise is to entertain a myth...or commit a fallacy (the false personification of a collectivity).” Framing analysis of policy implementation and performance as a matter of capacity focuses attention on a metaphor at the cost of abstracting away from the most salient features of the causal mechanisms that drive bureaucratic performance at both the organizational and political levels.

The focus on capacity also obscures the highly contingent nature of bureaucratic performance and policy implementation. “Capacity” may be a convenient shorthand for the complex array of factors that determines whether and how a particular policy is likely to be implemented in a specific case, but it achieves this convenience by abstracting away from the mechanisms that are critical for understanding and improving bureaucratic performance and policy implementation. Indeed, when discussing actual implementation of a specific policy by a specific bureaucracy, the abstract generality of state capacity quickly collapses into a much more complex set of explanations rooted in specific political, bureaucratic, and contextual contingencies. As such, the concept of state capacity is rarely analytically useful and is frequently misleading, and its unnuanced application to specific policies or contexts risks giving theoretical cover for not engaging with these important mechanisms and contingencies.

This critique of the concept of state capacity is not intended to denigrate the numerous excellent studies of bureaucratic effectiveness and reform that frame their analyses in terms of state capacity or capability but discuss it in a disaggregated way that recognizes the complexity and contingency of bureaucratic performance and political control (e.g. Soifer 2015, Andrews et al 2017, Centeno et al 2017). Rather, my argument is that in many cases

---

1 Other authors (e.g. Centeno et al 2017) have previously noted the distinction between state capacity and whether or how a political principle chooses to use that capacity. Throughout the article, I take this distinction for granted, and focus instead on the implications of multiple political principles for bureaucracies even after a policy decision has been made.
their nuanced analyses would be even more compelling were they not couched in state capacity. Moving beyond reliance on state capacity as an organizing concept for understanding state bureaucracies could foreground this nuance and open space for more useful analytical tools.

In response to the analytical inadequacies of state capacity, scholars should not simply substitute in another catch-all term that captures a similar underlying concept. To do so would be to focus on semantics rather than real conceptual issues. Instead, research on these topics should focus on actual actions rather than hypothetical potentials, for example by measuring retrospective performance rather than prospective capacity. Where the purpose of analysis is explicitly to predict how well a policy will be implemented, the analysis should explicitly engage with the complexity and uncertainty of policy implementation. These complexities are not well represented by a single unidimensional construct - whether retrospective or prospective - that is assumed to be fixed at the national, sub-national, or even organizational level. By asking specific questions about the likely outcomes of specific bureaucracies implementing specific policies, scholars and reformers alike can better understand and predict policy implementation and identify specific levers for meaningful improvement. This conceptual and analytical specificity would also feed directly into practical, problem-driven approaches to reform (Andrews et al 2017) and foster dialogue between comparative politics and other fields that study bureaucratic behavior and policy implementation, such as public administration.

While the lack of a direct conceptual replacement for state capacity complicates the work of scholars and reformers, engaging with these complexities is surely better than relying on a term whose main contribution is to obscure rather than clarify them. Doing so could also open up intellectual space for scholars to better understand successful bureaucracies in poor countries with generally weak states (Tendler 1997, Leonard 2010) as well as the numerous high-profile implementation failures in rich countries thought to have capable states (Dunleavy 1995), and begin to disaggregate theories of implementation and bureaucratic performance (Pritchett and Woolcock 2004; Pepinsky et al 2017). It would also connect more directly to the questions of efficiency and organizational dynamics that are the focus of much of the rich micro-level literature on bureaucratic performance from organizational economics and organization theory (Cyert and March 1963, Leibenstein 1966, Schein 1985, Gibbons and Henderson 2013), and to political science literature on legislative bargaining (Weingast and Marshall 1988) and political control of the bureaucracy (Whitford 2005). Ultimately, this process could lead to a clearer articulation of the connections between micro-level theories of bureaucracy and implementation and the macro-historical literature on state
formation from whence the concept of state capacity originated and spread. The remainder of this article proceeds as follows. Section 2 discusses how state capacity is defined and used as a measure of notional potential. Sections 3 and 4 show how the multiplicity of bureaucratic agents and political principals, respectively, undermine this view of the mechanisms of policy implementation and policy choice in government bureaucracies. Section 5 demonstrates that state capacity is a coarse way to understand the contingent nature of policy implementation and can be misleading with respect to reform, and suggests ways in which these critiques point towards more realistic and nuanced approaches. Section 6 concludes by discussing whether and how scholars should continue to use the term state capacity in light of these critiques.

2 The Concept of State Capacity

The term state capacity is used for a wide range of purposes by different authors, but this definitional diversity masks some key features that are common to its use in the governance literature on bureaucratic quality. Although a comprehensive review of these definitional and conceptual variations is beyond the scope of this article\(^2\), most uses in the governance literature are in the sense of what Mann (1984, 189) calls “infrastructural power”: “the capacity of the state to actually penetrate civil society, and to implement logistically political decisions throughout the realm.” Similarly, Skocpol (1985, 9) refers to “the ‘capacities’ of states to implement official goals, especially over the actual or potential opposition of powerful social groups or in the face of recalcitrant socioeconomic circumstances.” The sprawling research program that has followed can be divided into roughly three streams, following Soifer (2008, 232): 1) research focused on the “capabilities of the central state”; 2) research focused on the state’s “territorial reach”; and 3) research that emphasizes the “effects of the state on society”. My focus in this article is on the first of these, which emphasizes bureaucratic functioning and policy implementation, rather than the latter two, which focus more on the state’s monopoly on violence and the scope of the state.

Although the concept of state capacity originated in the macro-historical literature on state formation, it has been increasingly applied to questions of service delivery and policy implementation within the development and governance literatures. While different authors use different definitions, the common thread linking them is their emphasis on state capacity as a measure

of potential. For instance, Besley and Persson (2011, 6) define state capacity as “the institutional capability of the state to carry out various policies that deliver benefits and services to households and firms”, Andrews et al (2017, 95) define organizational capability as “the ability of an organization to equip, enable, and induce their agents to do the right thing at the right time to achieve a normative policy objective”, Kaufmann et al (2010, 4) refer to “the capacity of the government to effectively formulate and implement sound policies”, and Centeno et al (2017, 3) study “the organizational and bureaucratic ability to implement governing projects” (emphasis added throughout).

Since the ultimate goal is to explain or predict bureaucratic action - past, present, or future - this emphasis on measures of potential relies on implicit assumptions about the relationship between potential and action. In particular, potential and action can only be assumed equivalent to the extent that government bureaucracy can be modeled as a unitary agent implementing well-defined policy choices. Following the logic of constrained optimization, the bureaucracy is assumed to maximize the implementation of these policies subject to constraints of finite skills, knowledge, resources, and so on. This mental model of bureaucracy is analogous to simple economic models of firms’ production choices, in which a firm’s production possibility frontier represents possible solutions to the constrained optimization problem defined by its production function. Under these circumstances - when an organization can be thought of as maximizing output given a set of inputs - the metaphor of capacity is an accurate way to characterize governments’ ability to implement policy decisions. State capacity defines the frontier of combinations of public goods that could be produced, and politics is simply a matter of choosing a point along this frontier based on the political principal’s preferences and strategic calculations.

While most scholars of state capacity would recognize the reality of state bureaucracies to be more complicated than this simplistic characterization, the centrality of potential to the concept of state capacity is present even in its most nuanced treatments. For instance, Centeno et al (2017) distinguish organizational or state capacity from its political deployment, disaggregate state capacity into three dimensions and four indicators, and recognize the specificity of certain forms of state capacity while arguing against “generic notions of state capacity” (25). While these distinctions are all important and useful, the core of the concept nonetheless remains that state bureaucracies can usefully be conceived as having potential capacities that can be separated from actual actions, politics, and contextual specificities.

Does this conception of states as having bureaucratic capacity that can be politically deployed accurately depict the mechanisms of bureaucratic action
and behavior? The next two sections argue that it does not, even as a first approximation. Organizational performance and policy implementation is not a problem of constrained optimization by a unitary actor, but rather is a problem of coordination and commitment among multiple actors. Similarly, the most salient characteristic of political decisionmaking is its collective nature, which makes its outcomes unlikely to be a stable or efficient solution to a bureaucratic constrained optimization problem. While it is true that the analytical power of models comes from their simplifying assumptions, these assumptions should abstract away the least salient aspects of a problem in order to focus attention on the most important mechanisms. Conceiving of policy implementation and bureaucratic performance as problems of “capacity” achieves exactly the opposite.

3 Bureaucracies Are Collective Actors

States are composed of bureaucracies, or organizations, and organizations are collectivities of individuals. These individuals can be said to possess specific capacities, or (setting aside the many different types of skills and knowledge) some overall level of capacity. But there is no theoretical grounding for the assumption that the capacities of these individuals aggregate in any direct way to some collective organizational capacity. Indeed, a central theme of organization theory and organizational economics is that the collective nature of organizations introduces inefficiencies and complementarities, and thus organizations cannot be understood simply as the sum of their individual members. To adapt Cyert and March’s famous quote about the incoherence of the idea of organizational goals: people (i.e., individuals) have capacities; collectivities of people do not.

There are three sets of collective problems that undermine the analytical coherence of capacity as a concept for organizational analysis. First, there are simple problems of information and incentives that undermine individuals’ ability to collaborate efficiently within organizations. Second, there are problems of the allocation of individual capacity within and across organizations. Third, and most importantly, there are more complex problems of relational contracts and organizational culture that introduce the potential for multiple equilibria in organizational performance. While the first and second sets imply that there may be only a weak correlation between organizational performance and individual members’ capacities, the third set

\[\text{3While most scholars of state capacity would likely acknowledge that organizational capacity is not a simple aggregation of individual skills, it is nevertheless sometimes operationalized that way in practice (e.g. Gingerich 2013, Bersch \textit{et al} forthcoming).}\]
makes the stronger argument that lack of individual capacity is unlikely even to be a binding constraint for most organizations.\textsuperscript{4}

The first set of (relatively simple) problems stems from the idea that individual capacity centers on an individual’s ability to complete a given task, but in organizations these individuals face the additional challenge of coordinating their activities with each other. Indeed, the need to coordinate activities is a key reason why tasks are located within organizations in the first place, and is the subject of a large literature on team production (Marschak and Radner 1972). Cyert and March (1963) conceive of the organization as a process for making decisions when information is distributed, and Holmstrom (1982) identifies problems of moral hazard that arise from joint production in teams. Garicano and Rayo (2016, 138-9) neatly summarize the challenges imposed by the multiplicity of agents:

“Agents fail to act together because they do not want to (an incentive problem) or they do not know how to (a bounded-rationality problem). Incentive problems arise due to the presence of asymmetric information or imperfect commitment, which lead agents to act according to their own biases or preferences rather than in the interest of the organization (e.g., Holmstrom 1979; Shavell 1979). Bounded-rationality problems arise due to agents’ cognitive limitations and finite time, which means that even if they want to, agents cannot compute the solution to every problem, nor can they make themselves precisely understood by others…”

The stronger these incentive and bounded rationality problems are, the more that the organizations’ ability to resolve these problems will dominate the capacities of the individual agents in the determination of overall productivity. These problems are likely to be especially severe in the public sector, where outputs and outcomes are non-priced and often difficult to measure and managers’ ability to design and implement incentive schemes is typically constrained by statute and by politics (Wilson 1989).

The second set of issues, on the allocation of individual capacity, arises from the complementarities inherent in team production. If every worker in a team needs to perform a component of a task successfully in order for the overall task to be achieved, then the relationship between individual capacity

\textsuperscript{4}It is possible that in some cases higher individual capacity could even have a negative effect on overall organizational performance - for example, if higher capacity individuals are able to conduct “influence activities” inside organizations to advocate for their private interests at the expense of organizational objectives - although the scope for these is likely to be somewhat narrower than for the other types of collective problems discussed.
and team performance is multiplicative rather than separately additive. This is the premise of Kremer’s (1993) “O-ring” production process, inspired by the explosion of the immensely complex Challenger spacecraft due to the failure of a simple O-ring. In the most extreme cases, a team can only perform as well as its weakest worker.

These complementarities are pervasive in bureaucracies, particularly in the public sector. Many public sector outputs take the form of joint team production within or across organizations, as in the case where individuals from various units give inputs to different aspects of a permit decision or policy document. In addition, many public sector outputs require authorization from a sequence of individuals whose actions are informed not only by different mandates (e.g. finance, audit, public opinion, technical efficiency) but also different levels of individual capacity. Excessive bureaucratic delays, for example, often arise not from pervasive lethargy within government but from specific bottlenecks in the decision process. This can be illustrated with an example from the common problem of excessive delays in payment from government to infrastructure contractors. A Ghana Audit Service performance audit of the development of one large road project in Ghana (2013) found that the average period taken to pay each tranche to the contractor was over 200 days, compared to the 56 day payment period stipulated in the contract, resulting in delayed completion and over USD 2 million in interest payments. Of the 14 people that had to sign off on each payment certificate, the average delay was 15.8 days, but this masked immense variation: individual-level average delays ranged from 0.4 days to 86.7, with a standard deviation of 23.3. The implication of these types of joint or sequential production processes for bureaucracies is that increased individual capacity within one area of the organization - or one organization within the broader government - is unlikely to translate into a one-for-one improvement in overall performance, and may sometimes be entirely disconnected from it.

The third set of reasons for the disconnect between individual capacities and organizational performance centers on relational contracts and organizational culture within the organization, which can lead to multiple equilibria in organizational performance. These theories derive from the observation that many important aspects of organizational functioning are not formalizable and rely instead on informal understandings among members of the organization. The importance of non-formal practices and understandings in organizations has deep roots in organization theory and organizational economics, and starts from the observation that all real-world contracts are necessarily incomplete. This is particularly true of the employment contract and the management practices that specify the actions of managers and em-
Employees within this contract.\textsuperscript{5}

Contracts may be incomplete in two senses: future states of the world and the associated actions for each party to take may be impossible to fully anticipate and specify \textit{ex ante} (Simon 1951, Kreps 1996); and/or some value-creating aspects of parties’ actions may be unverifiable by third parties \textit{ex post}, and thus unenforceable (Hart and Moore 2008). This incompleteness implies that it is generally efficient for both parties to retain some level of discretion – indeed, as a practical matter it is almost impossible to completely eliminate discretion, and agent discretion has long been an important element of theories of management and public administration (Mintzberg 1983, Wilson 1989). However, discretion is a dual-edged sword: it can enhance efficiency for all parties, but can also be abused by actors for short-term private gain. The management of discretion is therefore both technical – in the specification of tasks, contingencies, and the design of incentives – but also relational - in that it requires building shared expectations, understandings, and norms over time. This accretion of shared understandings and processes over time is also a feature of Nelson and Winter’s (1982) influential work on routines in organizations, and creates the potential for substantial long-term divergences in performance among organizations (Chassang 2010).

Gibbons and Henderson (2013, 700-1) illustrate this idea with an example from Toyota:

“Workers were asked to be alert for potential opportunities for improvement and to be creative and innovative in their search for solutions. Successful participation in these teams relied on a set of behaviors that could not be articulated \textit{ex ante} or verified \textit{ex post}, and hence could not be rewarded through a formal contract. The use of the andon cord as a tool to improve the production process is a particularly compelling example. Every worker on the Toyota production line has access to an andon cord, a rope that can be pulled to alert a supervisor to the fact that the worker believes there is a problem on the line. Once the andon cord is pulled...the entire production line may be brought to a halt. This event is enormously costly. Giving individual workers the power to stop the line on the basis of something as nebulous as ‘whenever you see a problem’ implies...the existence of a relational contract

\textsuperscript{5}I focus on relational contracts that are within rather than between organizations. My discussion of relational contracts here is highly stylized and oriented towards building intuition about why performance might differ among organizations embedded in identical institutional, historical, socio-cultural, political, and task contexts. See Gibbons and Henderson (2013) for a concise formal exposition of these models.\textend{footnote}}
between workers and the firm. Any worker pulling the andon cord must exercise considerable judgment in identifying potential problems and must believe that he or she will not be penalized for potentially stopping the line. No formal contract can specify the conditions under which stopping the line is an appropriate thing to do.”

The relational character of the andon cord as management practice is emphasized by its failure when General Motors (GM) attempted to imitate Toyota and introduce the cord into its plants. Rather than serving as a tool for problem-solving and performance enhancement, the cord was undermined by mutual credibility problems: managers perceived that workers were simply pulling the cord whenever they wanted a break, while workers got yelled at when they pulled the cord (Helper and Henderson 2014).

Needless to say, employee discretion is also a salient feature of the public sector. Indeed, these relational aspects of management are likely to be even more important in public sector organizations than private sector ones, since the outputs of public sector organizations are often non-priced and/or difficult to measure (Wilson 1989, Prendergast 2003). The implication of the pervasive necessity for employee discretion in organizations is that all the formal aspects of management and policymaking that can be transported across organizations - standard operating procedures, remuneration and promotion schemes, descriptions of “best practices” - are not fully determinative of organizational performance. An implication of this is that two organizations that are identical in all formal aspects can exhibit major differences in performance of the same tasks, due to differences in how these informal, tacit understandings have developed among members of the organization.

There is considerable empirical evidence in support of the idea that ex ante identical organizations can exhibit large differences in performance. In developing country public sectors, the handful of quantitative studies that exist demonstrate large ranges of variation in performance within a given country’s government (Gingerich 2013; Rasul et al 2017), while a predominantly case study-based literature demonstrates the existence of “islands of excellence” - effective organizations - in otherwise weak states (Tendler 1997, Leonard 2010). In the private sector in OECD countries, numerous studies show large and persistent differences in productivity and management quality among organizations even within the same narrowly defined field (Gibbons and Henderson 2013), with evidence most heavily concentrated in the manufacturing sector (Bloom and Van Reenen 2007, Syverson 2011) but extending also to other fields such as hospitals (Bloom et al 2015, Carrera and Dunleavy 2013) and schools (Bloom et al 2014). These “persistent performance
differences” among ex ante identical organizations appear to be the norm, not the exception, within organizational fields (Gibbons and Henderson 2013). The potential for organizations to operate inefficiently has long been a key theme in the study of organizations, as theorists questioned models of firms as perfectly rational maximizers with concepts such as organizational slack (Cyert and March 1963), X-inefficiency (Leibenstein 1966), and organizational culture (Schein 1985). The potential for such variation in performance among public sector organizations is even greater, since there is no built-in mechanism for poorly performing government organizations to “exit” in the same way as inefficient firms.

The importance of relational contracts and organizational culture in organizations further weakens the usefulness of conceiving of government performance in terms of capacity. To the extent that these organization-specific relational factors matter for performance, improving performance becomes a question of shifting equilibria from an inefficient non-cooperative equilibrium to a more efficient cooperative one. Increasing the capacity of a single individual within a dysfunctional system may have little or no marginal effect on overall performance, and conversely, considerable gains in performance can be realized without any change in inputs.

This section has discussed how the collective nature of organizations as comprising multiple agents undermines the usefulness of capacity as an analytical device for understanding their performance or reform, as we are rarely likely to observe actual bureaucratic actions that are equivalent to the notional potential of the bureaucracy absent these inefficiencies imposed by its collective nature. The following section shifts to the level of the political control of state bureaucracies, to consider what implications the multiplicity of political principals has for the concept of state capacity.

4 Bureaucracies Have Multiple Principals

State capacity is defined as the ability of government bureaucracies “to implement logistically political decisions” (Mann 1984, 189; emphasis added). Similarly, Skocpol (1985, 9) discusses the “the ‘capacities’ of states to implement official goals”, and Besley and Persson (2011, 6) define state capacity as “the institutional capability of the state to carry out various policies that deliver benefits and services to households and firms”. If a state has a capable bureaucracy, the logic goes, then it should be able to effectively implement the government’s objectives, whatever they might be.

The implicit assumption in this is that governments actually have coherent and consistent goals that they can task an impartial bureaucracy to
implement without further political contestation. This is only true if a government’s goals are equivalent to those of a unitary actor - either because there is a clean separation between policy choice and policy implementation, so that all political disagreements are resolved at one stage and the resulting policy is implemented wholeheartedly, or because all decisions are taken by a dictator. Needless to say, neither of these conditions characterizes actually-existing governments. The concept of a clear distinction between policy choice and policy implementation has long been criticized in public administration (Fesler and Kettl 1991), and even within dictatorships numerous studies demonstrate that actual authority and decision power is dispersed across multiple actors (Bueno de Mesquita et al 2003). Much as Shepsle (1992) decried “legislative intent” to be an oxymoron by pointing out that “Congress is a ‘they’, not an ‘it’”, so too should scholars abandon the myth that the political process ordains coherent and consistent goals that bureaucracies could implement if only they were capable enough.

A more realistic approach would start from the recognition that government bureaucracies almost always have multiple principals (Wilson 1989, Dixit 1996), in the sense that their actions are directed and constrained by multiple actors, stakeholders, or objectives. These multiple principals are sometimes be embodied in formal institutions, as when bureaucracies are accountable to both the executive and a legislature (as well as to audit institutions, finance ministries, procurement authorities, etc.). Multiple principals can equally be understood in a less formal sense, in that bureaucracies are informally accountable to a broad range of stakeholders: organized interest groups, the media, “public opinion”, opposition political parties, professional bodies, service users, and so on. The very multiplicity of goals imposed on public sector organizations - effectiveness, transparency, impartiality, representativeness, etc. - also creates opportunities for new actors to direct or constrain the organization. As Wilson (1989, 131) observed, “Every constraint or contextual goal is the written affirmation of the claim of some external constituency.”

Finally, and to further complicate matters, the relevant multiple principals can even be internal to a bureaucracy. Cyert and March’s (1963, 205-6) observation about firms is even more applicable to the public sector:

“We have argued that the business firm is basically a coalition without a generally shared, consistent set of goals. Consequently, we cannot assume that a rational manager can treat the organization as a simple instrument in his dealings with the external world. Just as he needs to predict and attempt to manipulate the ‘external’ environment, he must predict and attempt to manip-
ulate his own firm. Indeed, our impression is that most actual managers devote much more time and energy to the problems of managing their coalition than they do to the problems of dealing with the outside world.”

This multiplicity of principals complicates the process of policy implementation, because each principal tries to influence how the policy is implemented throughout the implementation process. In other words, political contestation does not cease after the “decision” phase of policymaking. Whitford (2005, 45) describes the results of this “tug-of-war” on bureaucracies: “sequenced attempts by multiple, competing principals to obtain bureaucratic compliance can whiplash agencies as they implement policies in the field. For agencies, this shifting of gears - accelerating or decelerating as political overseers demand - has substantial importance for administration…” The challenges imposed on public managers by these competing, unstable, and collectively incoherent political demands, and their negative impact on efficiency and policy implementation, has also been extensively documented in qualitative literature (Pressman and Wildavsky 1973, Wilson 1989). More recently, Williams (2017) demonstrates that inability to resolve political disagreements among multiple principals could lead to public projects being abandoned mid-implementation despite this outcome being undesired by all actors, while Gulzar and Pasquale (2017) show that the presence of multiple principals undermines the efficiency of public service delivery in India.

Just as the fact of bureaucracies being collective rather than unitary actors suggests that “capacity” is a deeply flawed way to analyze organization’s ability to implement policy, so too does the characteristic of having multiple principals undermine the idea that official policy goals are coherent and stable objectives for these bureaucracies to aim at. As a result, the bureaucratic actions or performance we observe is likely to be very different from the notional potential of that bureaucracy. Understanding policy implementation as fundamentally about the “capacity” of public bureaucracies is therefore incomplete at best, and fundamentally misleading at worst. Yet if the concept of state capacity is a poor way to analyze bureaucratic performance and policy implementation both at the level of government bureaucracies and at the level of political control of these bureaucracies, then how should scholars analyze bureaucracies and implementation instead?
5 State Capacity, Policy Implementation, and Reform

State capacity’s popularity as a concept derives in large part from its practical usefulness, in two senses. First, state capacity claims to represent a state’s ability to implement policies - not just policies that are currently being implemented, but also hypothetical future policies. In this sense it is an essentially predictive concept: if a hypothetical policy were to be adopted, would it be implemented by the bureaucracy? Second, state capacity is useful because it creates a simple target for reform that is not only policy-neutral, but also unidimensional and apolitical. However, the critiques outlined in this article significantly undermine the validity of both of these uses. At the same time, they point towards more realistic and nuanced approaches to thinking about policy implementation and reform, as well as important new directions for research.

5.1 Capacity as Predictor of Implementation

The idea that state capacity is predictive of policy implementation in an abstract sense, divorced from specific policies and contextual contingencies, rests on two assumptions: 1) that policy implementation is primarily a matter of “capacity”; and 2) that this capacity is generalizable across various policies or tasks. With respect to the former, Section 3 argued that internal aspects of organizational functioning are dominated by issues of coordination and culture rather than individual capacity, while Section 4 argued that the policies bureaucracies are tasked to implement are often inconsistent or unstable and are subject to political interference in the implementation process, undermining the extent to which bureaucratic effectiveness alone determines policy implementation.

With respect to the generalizability of capacity, while it is possible to distinguish among better- and worse-performing bureaucracies both within countries (Rasul et al 2017) and across them (Kaufmann et al 2010), these overall levels do not preclude major deviations: there are numerous high-profile implementation failures in rich countries with strong bureaucracies (Dunleavy 1995), as well as implementation successes in countries with less effective states (Tendler 1997, Leonard 2010). Moreover, the extent to which effectiveness in one task, policy domain, or aspect of the organization translates into effectiveness in hypothetical new policy domains is debatable. While some core organizational functions (audit, accounting, human resources, monitoring) are likely to enhance the performance of whatever policy a bureau-
cracy might be asked to implement, other public sector functions are clearly not like this: running a clinic, enforcing financial regulations, and building roads obviously require different sets of technical skills and organizational processes, which is why they are carried out by different organizations. What little empirical evidence that exists reinforces these doubts: for example, Rasul and Rogger (2016) and Rasul et al (2017) demonstrate in Nigeria and Ghana’s governments that bureaucracies’ optimal management practices vary across project types, implying that (since most bureaucracies implement multiple types of projects or tasks) that there is not a single set of management practices that would be optimal for all bureaucracies to adopt.

The importance of contextual specificities and contingencies becomes even more clear once abstract discussions of state or organizational capacity are narrowed to specific policy questions. Consider a hypothetical example from an area - tax collection - that is central to theories of state capacity (e.g. Besley and Persson 2009, Harbers 2015). Suppose one wanted to predict whether the tax authorities of a given country - Kenya, the United States, or Thailand, say - could increase tax collections by ten percent. To answer this question would require engaging with a broad range of contextual factors - resources, time pressures, the design and complexity of the policy instrument, political constellations and circumstances, and professional incentives - among which the past performance of the tax agency is only one. Would the agency be given additional resources to achieve this increase? Over what timeline? Would the political environment even permit this policy to be adopted, and to make any necessary statutory changes? What interest groups would be able to influence the policy’s implementation after it was adopted? Is this increase to be achieved by an easy-to-administer increase in rates, or a more challenging expansion of the tax base? What incentive would the leadership, middle managers, and frontline workers of the agency have to achieve this goal? What coordination issues among agency staff would have to be resolved, and would the organization’s existing culture support or hinder this? While an especially high-performing agency might be able to resolve some of these implied challenges better than a poorly performing agency, the ability of even the best agency to achieve that goal would depend on the answer to these questions, many of which are out of the agency’s control.

The importance of contextual contingencies also emerges clearly from retrospective analyses of policy implementation. Centeno et al’s (2017, 19-20) discussion “of the presence of state capacity needed to successfully achieve land reforms” is illustrative. Successful land reforms:

...have followed either successful social revolutions, as in China,
Vietnam, or Cuba, or have occurred under foreign military occupation aimed at thwarting communism, as in South Korea or Taiwan during the early 1950s. In all such cases, states had pressing reasons to redistribute the land: In addition to ideological motivations in communist cases, peasants often provided social support to revolutionary leaders, and these leaders, in turn, needed to destroy the landed gentry as part of their state-building strategy; and similarly, in the anticommunist cases, peasant demands for land and attraction to communism needed to be neutralized as part of building a state that could withstand the communist threat. Beyond elite goals, the success of land reforms also underlines the type of state capacity needed to implement policies that challenge powerful social interests. States in all of these cases were simultaneously centralized and decentralized. A well-organized communist party or an occupation army linked central power in capital cities to the vast peripheries. Key central political decisions—such as to redistribute land—could then travel down the political hierarchy of the party (in the communist cases) or the occupying military (in the anticommunist cases).

The analysis is detailed and nuanced, foregrounding unique political circumstances and motivations, state-society linkages, and particular organizational forms and state structures. The authors also note that “[t]he ‘redistributive coalition’ that pursues land reforms tends to be very different than the ‘growth coalition’ that steers developmental states…” (20). Even for the two (quite broad) policy goals of land reform and economic growth that have both been strongly linked to state capacity, the concept of generalized state capacity provides little analytical leverage in light of these other highly contingent circumstances influencing bureaucratic action and performance. Despite Centeno et al.’s (2017) careful and nuanced treatment, the concept of state capacity on its own appears a blunt tool for analyzing the implementation of land reforms.

The case of the failed launch of the healthcare.gov website in the United States provides an even sharper contrast between the abstract generality of state capacity and the highly contingent nature of actual policy implementation. The US federal government would be judged as high capacity by any measure and the website was delivering on a top political priority, yet the launch was a dismal failure. This failure was caused by a combination of technical complexity, poor project management, unrealistic politically driven timelines, ineffectual risk analysis, and poor coordination among stakeholders (Anthopolous et al 2016). These can be understood in the context of
the multiple agent and multiple principal theories of bureaucracy discussed in Sections 3 and 4. Although the government had many individuals with appropriate technical skills (and had the resources to hire many others), implementation was characterized by the inefficient allocation of this individual capacity across government, by coordination failures, and by hierarchical working norms that were inappropriate for such a complex project - problems imposed by the collective nature of bureaucracies. Similarly, fragmented authority across within the government meant that “key decisions were often delayed, guidance to contractors was inconsistent, and nobody was truly in charge. Government employees appear to have concealed critical information from each other…” (Thompson 2013) - all manifestations of the multiple principal character of public sector organizations.

State capacity is thus a coarse and unhelpful concept for scholars to use to analyze bureaucracies' policy implementation outcomes. One response to this is to focus research and measurement on actual bureaucratic actions, and on retrospective performance rather than prospective capacity. When what is needed are analytical tools with which to analyze prospective implementation outcomes, however, analysts may have no choice but to embrace the contingency of implementation on the specificities of policies, policies, contexts, and organizations. While this is inconvenient in that it precludes reliance on uni- or low-dimensional abstract measures like state capacity, greater attention to such contingencies may simply be the price for improved validity and analytical leverage.

5.2 Capacity as Target for Reform

To what extent does state capacity - and its practice-based manifestation, capacity building - present a useful target for reform? This question is of significant practical consequence, as “a quarter of the US $55 billion of total Overseas Development Assistance is accounted for by support for capacity building” (World Bank 2016, 1). At the same time, the results of donor-supported capacity building programs are widely viewed as disappointing and lagging behind other areas of development assistance (OECD 2006, World Bank 2016). The remainder of this section argues that the reasons for this ineffectiveness can be located in the conceptual limitations discussed above - as can steps towards a more productive approach.

Reforms conceived as capacity-building began in the 1970s, with donor-funded technical assistance programs that focused mainly on improving individual skills (Teskey 2012). Yet the relationship between individual skill development and improved performance in collective actors such as bureaucracies, for which successful policy implementation is highly contingent and
often requires specific skills, is far from direct. Not only does investing in skills that might only potentially be used create waste, but performance and policy implementation are determined in large part by unformalizable and relational aspects of organizational functioning which can only be put in place and improved through actual practice. Investing in potential capacities is therefore likely to be ineffective unless they are actualized. Similarly, the productivity of individual skills and capacities depends a great deal on their allocation, coordination, and incentives, as donor evaluations of the failures of individual-focused capacity building came to recognize (OECD 2006).

In response to these perceived failings, donors broadened the definition of capacity to include organizational and institutional factors over the course of the 1980s and 1990s (Teskey 2012). This led to understandings of capacity that were so broad as to encompass virtually anything to do with government performance, as exemplified by the OECD Development Assistance Committee’s (2006, 12) definition of capacity as “the ability of people, organisations and society as a whole to manage their affairs successfully.” Yet despite the awareness of the importance of organizational and institutional factors in bureaucracies’ performance, the most commonly used approach to capacity building in practice has remained individualized skill development through trainings and workshops, despite the skepticism of these very practitioners about the usefulness of such approaches (USAID 2017).

This paper’s critique of the usefulness of capacity as a framework for understanding bureaucratic performance and policy implementation points toward two ways in which reforms might be more effectively designed, corresponding to the two key features of bureaucracies (multiple agents, multiple principals) noted in Sections 3 and 4. First, programs should focus on improving performance rather than capacity. If important aspects of organizational functioning such as coordination and relational practices can only be developed through action, then working to improve these in the abstract - divorced from the implementation of specific policies - will be ineffective. This implies a greater focus on improving team dynamics and organizational culture through learning-by-doing - the types of activities which in the private sector, as Teskey (2012, 1) astutely notes, are simply called “business management” rather than capacity building. This emphasis on performance over capacity has recently been adopted in measurement and evaluation by USAID (2017, 5): “Capacity is a form of potential; it is not visible until it is used. Therefore, performance is the key consideration in determining whether capacity has changed.” A focus on performance also complements problem-driven and adaptive approaches such as Andrews et al (2017) - although the authors refer to it as capability building - that emphasize collective applied problem-solving and iterative improvement over time.
Second, understanding the implications of bureaucracies’ multiple principals can also contribute to more precise analysis of the ways in which politics might affect policy implementation - both negative and positive - and improve the design of institutions meant to ameliorate these effects. Conceiving of bureaucracies as being pulled among multiple political principals makes reform more difficult in some ways, by increasing the number of constraints on bureaucracies - a perspective taken by applied literature on political economy analysis in aid delivery (DFID 2009). Yet it also emphasizes the potential scope for bureaucratic autonomy to have significant positive effects on the policy process (Miller and Whitford 2016), and the importance of finding creative ways for the leaders of these bureaucracies to carve out autonomy (Carpenter 2001). Similarly, understanding the importance and implications of multiple principals can have significant implications for institutional design of accountability systems for bureaucrats (Gulzar and Pasquale 2017) and delivery mechanisms for aid and inter-governmental transfers (Williams 2017). Recognizing the limitations of state capacity for thinking about bureaucratic performance, policy implementation, and reform thus opens up avenues for future research and policy innovation that are both theoretically grounded and practical.

6 Conclusion

This article has argued that, despite its ubiquity in the literature, state capacity misrepresents both the mechanisms through which bureaucracies act, and the highly contingent nature of these actions. The metaphor of “capacity” obscures the salient fact that bureaucracies are collective actors operating under the direction of multiple political principals. As a result of this multiplicity of agents and principals, most failures of service delivery or policy implementation are not primarily due to the capacity of the individuals that comprise these bureaucracies, but rather to problems of coordination, collective action, and collective choice. Rather than basing analysis of policy implementation and reform on the idea that bureaucracies can be effectively characterized by some generalized notion of capacity divorced from the implementation of specific policies, researchers and policymakers should instead focus on performance rather than capacity, and - where it is necessary to predict implementation outcomes prospectively - analyze specific challenges to implementing specific policies in specific bureaucracies.

This is not to argue that state capacity should be banished altogether from the scholarly lexicon. While state capacity as a concept lacks firm foundations in theory or empirics on bureaucratic behavior, this lack of nu-
ance and realism matters more for some research questions than others. This article has focused mainly on questions of processes of policy implementation and reform for which nuance and detail are essential and so the costs of abstraction outweigh the benefits - conditions which characterize most considerations of specific policies in specific contexts. However, much of the research on state capacity focuses on questions of long-term historical development of the state (Besley and Persson 2011) or on broad cross-country comparisons (D’Arcy and Nistotskaya 2016) - indeed, the concept originated in these literatures (Mann 1984), and only later spread to research on contemporary states, specific policy domains, and reform. When taking such a bird’s-eye view of bureaucracy, the conceptual abstraction of state capacity might be appropriate as a shorthand way to refer to the complexities of bureaucratic effectiveness and policy implementation. Similarly, for some research questions where state capacity is a mediating variable rather than the question of interest (e.g. Englehart 2009), state capacity’s abstraction seems less problematic - although even in these cases, the mechanisms and contingencies highlighted in this paper could still be consequential.

Research on the development and performance of government bureaucracies has made impressive progress in the three decades since works like Mann (1984) and Skocpol (1985) began to make it a key research question for social science, and much of this progress has been made under the banner of “state capacity”. To further advance, however, research on bureaucracies will require a more nuanced and theoretically grounded approach to understanding bureaucratic performance, policy implementation, and reform. Fortunately, many of the theoretical and methodological building blocks for this approach already exist within political science, organization theory, and organizational economics, and much can be drawn from the more disaggregated and empirically careful studies of state capacity that already exist. Unfortunately, even the most sophisticated models only begin to make sense of the complexity of state bureaucracies, as numerous reformers have discovered the hard way. Integrating these insights into theory and empirics on policy implementation and reform, and working with policymakers to test them in the real world, represents a rich research agenda.

Even in its most abstract treatments, however, the collective nature of bureaucracies and their multiple principals can be consequential. For instance, Besley and Persson (2009) refer to rulers’ decisions to “invest” in state capacity. Yet as discussed above, the translation of political desires into bureaucratic output is far from the simple, deterministic function that Besley and Persson describe, and could vary in form according to some of the same political economy characteristics that drive their model.
References


Skocpol, Theda. 1985. “Bringing the State Back in: Strategies of Analysis in Current Research.” In Evans, Peter, Dietrich Rueschmeyer, and Theda Skocpol (eds), Bringing the State Back In, Cambridge: Cambridge University Press: 3-43.


