

RESEARCH ARTICLE

Beyond state capacity: bureaucratic performance, policy implementation and reform

Martin J. Williams 

Blavatnik School of Government, Radcliffe Observatory Quarter, University of Oxford, Oxford OX2 6GG, UK
Corresponding author. Email: martin.williams@bsg.ox.ac.uk

(Received 8 December 2019; revised 24 September 2020; accepted 25 September 2020; first published online 28 October 2020)

Abstract

There is a broad consensus that state capacity is central to economic and institutional development. But while the concept originated as a tool for macro-historical and comparative analysis, its success has led the term ‘capacity’ to become a default metaphor for discussing the quality of government bureaucracies. This paper discusses the limitations to conceiving of narrower questions of bureaucratic performance and policy implementation through the lens of the broad, aggregate concept of capacity. Whereas capacity refers to bureaucracies’ hypothetical potential, this usually differs from their actual actions due to internal information and incentive problems created by bureaucracies’ collective nature, and the constraints and uncertainty imposed by their multiple political principals. Capacity is a convenient shorthand term and is appropriate for some purposes, but it achieves this convenience by abstracting away from the mechanisms that determine bureaucratic performance and policy implementation. To advance the study of bureaucratic quality, researchers should seek to understand the implications of bureaucracies’ collective nature, engage with contextual specificity and contingency in policy implementation, and focus measurement and reform efforts more towards actual performance than hypothetical capacity.

Key words: State capacity; government performance; bureaucracy; policy implementation

JEL Codes: H8; H11

1. Introduction

State capacity – the ability of the state bureaucracy to implement government’s policy choices – has become one of the most influential concepts in research on international development. The sprawling research programme originated in the effort to push analysis beyond analysis of politics and policy decisions into the realm of logistics, power and implementation of these decisions (Mann, 1984; Skocpol, 1985; Tilly, 1975), and intersected with the Northian effort to understand the role of the state in institutional development (Acemoglu and Robinson, 2012; Acemoglu *et al.*, 2001; Besley and Persson, 2009; Evans and Rauch, 1999; North *et al.*, 2009). The key theoretical and empirical questions were macro-historical: explaining why and how strong security and administrative bureaucracies developed in some states but not in others. While an active literature still debates the conceptualisation, measurement, historical determinants and consequences of state capacity’s development across countries and regions, a measure of the success of this literature is that few scholars would now contest that strong, capable states are central to processes of long-run development (e.g. Besley and Persson, 2011; Centeno *et al.*, 2017; Chuaire *et al.*, 2017; Englehart, 2009; Hanson and Sigman, 2019; Harbers, 2015; Hendrix, 2010; Lee *et al.*, 2014; Salter and Young, 2019; Soifer, 2008).

At the same time as this research programme has been examining the relationship between state capacity and socioeconomic outcomes at an aggregate level, a related set of concerns around the quality of government bureaucracies has become increasingly prominent in more narrowly focused

research on particular bureaucracies or policies and in development practice. These efforts often appeal to the theoretical apparatus and terminology of capacity, perhaps in part due to the success and intellectual influence of the macro-historical and cross-country state capacity literature and the concept's malleability. For example, institutional reform is often referred to as a matter of building capacity or capability (Andrews *et al.*, 2017; Teskey, 2012), impact evaluations are framed as evaluating the effects of state capacity through specific programmes (Muralidharan *et al.*, 2016), research on the implementation of particular laws or policies is analysed as a matter of capacity (e.g. Baker, 2009; Hills, 2007), and government organisations' capacity is measured by aggregating personnel indicators (Bersch *et al.*, 2016; Gingerich, 2013). How consequential is this shift from using capacity as a macro-historical concept to examine the relationships between aggregate outcomes and broad measures of state quality to using capacity as a theoretical lens for narrower questions of bureaucratic performance, policy implementation and reform?

This paper argues that the concept of bureaucratic capacity becomes less useful for understanding government bureaucracies the farther one moves away from the broad, aggregate abstraction of the original macro-historical concept towards the discussion of specific bureaucracies, policies or reforms. While the metaphor of bureaucratic capacity is intuitive and appealing, when applied to specific organisations or policies it misrepresents the mechanisms of bureaucratic performance and policy implementation and obscures the contingency of performance and implementation on the details 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. Organisations 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 Rayo, 2016) 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 (Shepsle, 1992; Wilson, 1989).¹

While individuals may thus be said to have specific capacities, conceiving of organisations as having capacities obscures perhaps the most salient characteristic of organisations: that they are collective actors. As Shepsle (1992: 339) writes in his classic polemic (entitled 'Congress Is a "They", Not an "It"') against the concept of legislative intent: 'To claim otherwise is to entertain a myth ... or commit a fallacy (the false personification of a collectivity)'. Analysing issues such as bureaucratic performance, policy implementation and reform through the lens of capacity also obscures their highly contingent nature. Capacity is 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. 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 organisational and political levels. While such abstraction may be appropriate for some purposes, for other purposes it risks obscuring the very phenomena it aims to analyse.

How should scholars and practitioners approach these questions, then? The answer is not to simply substitute in another catch-all term to capture a similar underlying concept; to do so would be to focus on semantics rather than real conceptual issues. Instead, this paper suggests three (non-exclusive) approaches. First, research on organisational performance and reform should explicitly engage with

¹Other 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 paper, I take this distinction for granted, and focus instead on the implications of multiple political principals for bureaucracies even after a policy decision has been made.

the implications of bureaucracies being collective actors under multiple principals. Second, analysis must engage directly with the contingency and specificity of policy implementation, which is not well represented by a single unidimensional construct that is assumed to be fixed at the national, sub-national or even organisational level. Finally, work on these topics should carefully distinguish between actual actions and hypothetical potentials: whereas retrospective performance can be measured, discussions of prospective capacity are inherently speculative.

Although this paper critiques much of the conceptual slippage that has been associated with state capacity, it does not call for the concept's abandonment or denigrate the numerous excellent studies that have advanced our understanding of the historical and comparative development of state capacity (e.g. Andrews *et al.*, 2017; Mkandawire *et al.*, 2017; Bologna Pavlik and Young, 2020; Centeno *et al.*, 2017; Chuaire *et al.*, 2017; Salter and Young, 2018, 2019; Soifer, 2008). Rather, the paper calls attention to a common trade-off for theory: a concept designed for analysis at high levels of abstraction is likely to be less suited for less abstract and more specific questions (and vice versa). The same aggregation and acontextuality that make state capacity such a powerful concept for studying the types of questions for which it was intended also inherently limit its application to the narrower questions of bureaucratic performance, policy implementation and reform to which it has been increasingly applied.

Recognising this limitation opens space for developing and connecting other theoretical approaches to these issues. For example, engaging with the complexities of policy implementation can help scholars better understand successful bureaucracies in poor countries with generally weak states (Leonard, 2010; McDonnell, 2017; 2020; Tandler, 1997) 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 (Pepinsky *et al.*, 2017). It would also connect more directly to the questions of efficiency and organisational dynamics that are the focus of much of the rich micro-level literature on bureaucratic performance from organisational economics and organisation theory (Cyert and March, 1963; Gibbons and Henderson, 2013; Leibenstein, 1966; Schein, 1985), 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, organisations and policy implementation, and the macro-historical literature on state formation and institutional development from which the concept of state capacity originated and spread.

The remainder of this paper proceeds as follows. Section 2 discusses the origins and development of state capacity in the macro-historical literature, and section 3 describes the concept's slippage into narrower, more applied areas of research and practice. Sections 4 and 5 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 6 discusses three ways in which scholars and policymakers can respond to these critiques, and section 7 concludes.

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 paper,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 programme 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

²See Soifer (2008), Cingolani (2013), Centeno *et al.* (2017), and Hanson and Sigman (2019) for useful reviews.

the state's 'territorial reach' and (3) research that emphasises the 'effects of the state on society'. My focus in this paper is on the first of these, since this meaning is most relevant for the issues of policy implementation and bureaucratic performance that are the main subject of this paper. Within this, Hanson and Sigman (2019) make an important distinction between narrow questions of bureaucratic capacity or quality and broader questions of political institutions and good governance; I follow Hanson and Sigman in focusing my discussion on the former.

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', Kaufmann *et al.* (2010: 4) refer to 'the capacity of the government to effectively formulate and implement sound policies', Centeno *et al.* (2017: 3) study 'the organisational and bureaucratic *ability* to implement governing projects', and Hanson and Sigman (2019: 2) define state capacity as 'the *ability* of state institutions to effectively implement official goals' (emphases added throughout).

For analysts who seek 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 modelled as a unitary agent implementing well-defined policy choices. Following the logic of constrained optimisation, the bureaucracy is assumed to maximise 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 optimisation problem defined by its production function. Under these circumstances – when an organisation can be thought of as maximising output given a set of inputs – the metaphor of capacity is an accurate way to characterise 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 recognise the reality of state bureaucracies to be more complicated than this simplistic characterisation, 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 organisational or state capacity from its political deployment, disaggregate state capacity into three dimensions and four indicators and recognise 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.

3. Capacity in applied and practice-oriented research

At the same time as state capacity was becoming a central issue in the study of long-term development, so too was capacity becoming an increasingly common analytical framework and theoretical reference point for scholars of bureaucratic performance and practitioners of institutional reform. Understanding bureaucratic (in)action as a matter of capacity was also convenient for these more applied purposes, in large part due to the concept's malleability. Much like state capacity, capacity building as a concept has frequently been criticised for weak theoretical underpinnings, a range of definitions and a tendency towards integrating a wide range of phenomena into a single term (Baser and Morgan, 2008; Bockstael, 2017). Capacity is used with reference to individuals, organisations, communities, systems and nations alike (Ubels *et al.*, 2016). Brinkerhoff and Morgan write that

'Exploring capacity can have an Alice-in-Wonderland feel: different definitions and models inhabit disjunctive realities where underlying assumptions are neither obvious nor transferrable. Like Alice, we wander through these worlds in varying states of befuddlement or irritation. As Morgan (2003: 1, in Brinkerhoff and Morgan [2010]) notes, the concept of capacity "seems to exist somewhere in a nether world between individual training and national development"' (2010: 2). While not attempting a full survey of the literature on capacity building or policy implementation,³ this section discusses the prevalence of the concept of capacity not only in contemporary development practice but also in research on government performance and policy implementation, showing that it shares with the macro-historical literature on state capacity an emphasis on *potential* bureaucratic actions.

Certainly in financial terms, capacity building has become central to contemporary development practice, 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), and Brinkerhoff and Morgan note that 'Attention to capacity and capacity development (CD) has endured since the birth of international assistance' (2010: 2). Reforms conceived as capacity-building began in earnest in the 1970s, with donor-funded technical assistance programmes that focused mainly on improving individual skills (Teskey, 2012). Yet, these individual-focused programmes were widely perceived to have failed in their impact (OECD, 2006). In response to these perceived failings, donors broadened the definition of capacity to include organisational and institutional factors over the course of the 1980s and 1990s (Teskey, 2012).

This has led to understandings of capacity or capability that are so broad as to encompass virtually anything government or an organisation does: 'the *ability* of people, organisations and society as a whole to manage their affairs successfully' (OECD, 2006: 12); the '*potential* to perform' (Horton *et al.*, 2003: 18); 'the *ability* of a human system to perform, sustain itself and self-renew' (Ubels *et al.*, 2010); and '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' (Andrews *et al.*, 2017: 95; emphasis added throughout). Despite the breadth of these definitions – and as with state capacity in the macro-historical literature – the common thread of these definitions is their emphasis on the *potential* of bureaucracies to achieve certain objectives.

The use of capacity as a framing device and organising concept extends beyond development practice into academic research. For example, Briebea (2018) studies the evolution of Argentina and Chile's performance on infant and maternal mortality indicators, and finds that 'investments in state capacity' – used synonymously with development of the health system – were key to Chile's superior performance. An *et al.* (2017) examine how various 'capacity factors' affect the delivery of urban infrastructure in India. Bruszt and Campos (2019) couch their sophisticated study of institutional reform and economic integration in terms of state capacity. The capacity metaphor also extends beyond government organisations, for instance in the large literature on the development of 'community capacity' to resolve governance challenges (e.g. Moreno *et al.*, 2017).

Capacity's appeal as a conceptual framework for applied policy and academic research derives in large part from its practical and political usefulness, in three senses. First, unnuanced readings of capacity see it as a way to compress many potential dimensions of organisational performance into a uni-dimensional concept that an organisation can simply have more or less of. Second, capacity is defined as 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? This not only corresponds to the practical interests of bureaucrats and donors, but also neutralises political disagreements about what government should be doing. Third, capacity as a concept is useful because it creates a simple *target for reform* that is policy-neutral and apolitical. This makes it possible to discuss making changes to state structures and processes without being seen to be intervening in political arenas.

³See Baser and Morgan (2008) for a review.

These features perhaps explain the particular attractiveness of the term in the field of international development, where international donors are especially important and many researchers view themselves as external to the context (Bertelli *et al.*, forthcoming). The context-free nature of capacity as a term appeals to donors who operate across (and thus want to compare) multiple countries, and where headquarters offices may lack contextual information on particular countries or organisations (Honig, 2018). Similarly, donors and other external actors are typically keen to be seen as apolitical; construing bureaucratic reforms as being about improving governments' ability to do whatever it wants has the rhetorical effect of minimising perceived threats to sovereignty.

Given this appeal it is not surprising that policymakers and academics alike have adopted the concept so readily. But does this conception of states as having bureaucratic capacity that can be politically deployed accurately depict the mechanisms of bureaucratic action and behaviour?

4. Bureaucracies are collective actors

States are composed of bureaucracies, or organisations, and organisations 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 these individual capabilities aggregate into collective organisational action through mechanisms that are often complex and indirect. Indeed, a central theme of organisation theory and organisational economics is that the collective nature of organisations introduces inefficiencies and complementarities, and thus organisations can amount to more or less than the sum of their individual members. This is exemplified by Cyert and March's quip that 'People (i.e. individuals) have goals; collectivities of people do not' (1963: 30). Similarly, while individuals can be said to have capacities, collectivities of people do not – or at least not in a simple sense.⁴ Of course, much of the literature on state capacity understands it as a collective phenomenon rather than a simple aggregation of individual skills – although it is sometimes operationalised that way in measuring bureaucratic quality (e.g. Bersch *et al.*, 2016; Gingerich, 2013). But unpacking the implications of organisations as collective actors should lead us to be skeptical about the usefulness of the capacity metaphor for understanding organisational action, as the remainder of this section discusses.⁵

The collective problems that undermine both individuals' ability to collaborate efficiently within organisations and the analytical coherence of capacity as a concept for organisational analysis can be understood in terms of problems of information and of incentives. While such challenges are of course pervasive and broad in scope, this section highlights two such sets of issues that are particularly relevant for the topic of bureaucratic capacity. First, there are problems of the allocation of individual capacity within and across organisations. Second, and more importantly, there are more complex problems of relational contracts and organisational culture that introduce the potential for multiple equilibria in organisational performance. While this first set of challenges (as well as the presence of these information and incentive problems more generally) implies that there may be only a weak correlation between organisational performance and individual members' capacities, the second set makes the stronger argument that lack of individual capacity is unlikely even to be a binding constraint for most organisations.

⁴While there are debates within sociology and philosophy about when corporate actors might be said to have agency (c.f. Coleman, 1982; List and Pettit, 2011; Tuomela, 1991), these consider a rather more nuanced view of collective agency or capability than is generally deployed in discussions of bureaucratic capacity.

⁵My focus in this section is on collective action in the sense of the complexities of getting many agents to act together *within* a given bureaucracy in order to produce organisational action, rather than the alternative senses of bureaucracies solving collective action problems by producing public goods or of the collective action problems inherent in getting political principals to invest in building effective states. Both these latter senses have been extensively explored in the state capacity literature (e.g. Acemoglu *et al.*, 2015; Besley and Persson, 2009; Dincecco and Katz, 2014; Johnson and Koyama, 2017), but are distinct from my focus in this section.

Individual capacity centres on an individual's ability to complete a given task, but in organisations these individuals face the additional challenge of coordinating their activities with each other. Garicano and Rayo (2016: 138–139) neatly summarise 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 organisation (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 organisations' 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).

Within this broad scope, the more specific set of issues regarding the allocation of individual capacity arise 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 and team performance is multiplicative rather than separately additive. These complementarities are pervasive in bureaucracies (Alchian and Demsetz, 1972), particularly in the public sector. Many outputs take the form of joint team production within or across organisations, as when individuals from various units give inputs to different aspects of a permit decision or policy document. In addition, many public sector outputs require authorisation from a sequence of individuals whose actions are informed not only by different mandates but also different levels of individual capacity. The implication of these types of joint or sequential production processes for bureaucracies is that increased individual capacity within one area of the organisation – or one organisation 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 second more specific set of reasons for the disconnect between individual capacities and organisational performance centres on *relational contracts* and *organisational culture* within the organisation, which can lead to multiple equilibria in organisational performance. These theories derive from the observation that many important aspects of organisational functioning are not formalisable and rely instead on informal understandings among members of the organisation (Gibbons and Henderson, 2013). This incompleteness implies the need for agents to retain some level of discretion, but 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 organisations, and creates the potential for substantial long-term divergences in performance among organisations.

Needless to say, employee discretion is a salient feature of the public sector. Indeed, these relational aspects of management are likely to be even more important in public sector organisations than private sector ones, since the outputs of public sector organisations are often non-priced and/or difficult to measure (Prendergast, 2003; Wilson, 1989). The implication of the pervasive necessity for employee discretion in organisations is that all the formal aspects of management and policymaking that can be

transported across organisations – standard operating procedures, remuneration and promotion schemes, descriptions of ‘best practices’ – are not fully determinative of organisational performance. An implication of this is that two organisations 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 organisation.

There is considerable empirical evidence in support of the idea that *ex ante* identical organisations 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.*, forthcoming), while a predominantly case study-based literature demonstrates the existence of ‘islands of excellence’ – effective organisations – in otherwise weak states (Leonard, 2010; McDonnell, 2017, 2020; Tendler, 1997). Numerous studies of private sector firms show large and persistent differences in productivity and management quality among organisations even within the same narrowly defined field (Gibbons and Henderson, 2013), as well as in other fields such as hospitals (Carrera and Dunleavy, 2010) and schools (Bloom *et al.*, 2014) and schools (Leaver *et al.*, 2019). These ‘persistent performance differences’ among organisations appear to be the norm, not the exception, within organisational fields (Gibbons and Henderson, 2013). The potential for organisations to operate inefficiently has long been a key theme in the study of organisations, as theorists questioned models of firms as perfectly rational maximisers with concepts such as organisational slack (Cyert and March, 1963), X-inefficiency (Leibenstein, 1966) and organisational culture (Schein, 1985). The potential for such variation in performance among public sector organisations is even greater, since there is no built-in mechanism for poorly performing government organisations to ‘exit’ in the same way as inefficient firms.⁶

The importance of relational contracts and organisational culture in organisations further weakens the usefulness of conceiving of government performance in terms of capacity. To the extent that these organisation-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. Capacity may be a coherent way to understand individual actions given a set of incentives, but treating organisations as having collective capacity abstracts away from the most salient mechanisms that drive individual bureaucratic actions and collective organisational performance. This is not to claim that a bureaucracy’s hypothetical capacity is orthogonal to its actual performance, but rather that approaching such questions through the lens of capacity risks obscuring the very mechanisms it wishes to understand – an issue to which this paper returns in section 6.

5. Bureaucracies have multiple principals

State capacity is defined as the ability of government bureaucracies ‘to implement logistically political decisions’ (Mann, 1984: 189). 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.

An 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 characterises actually-existing governments. 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

⁶I am grateful to Patrick Dunleavy for this point.

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* (Dixit, 1996; Wilson, 1989), in the sense that their actions are directed and constrained by multiple actors, stakeholders or objectives. These multiple principals are sometimes 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: organised 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 organisations – effectiveness, transparency, impartiality, representativeness, etc. – also creates opportunities for new actors to direct or constrain the organisation. 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: 205–206) 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 manipulate 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).

Just as the fact of bureaucracies being collective rather than unitary actors suggests that capacity is a flawed way to analyse organisation’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 are likely to be very different from the notional potential of that bureaucracy. As in the previous section, the claim here is not that notional capacity is unrelated to actual actions, but rather that the multiplicity of principals makes understanding policy implementation as a question of the capacity of public bureaucracies an incomplete way to analyse what bureaucracies do or do not do. Yet, if the concept of capacity is an inherently limited lens on 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 analyse bureaucracies and implementation instead?

6. Discussion: beyond capacity

It would be convenient if the appropriate response to the limitations of the capacity framework was simply to adopt a different term. However, the limitations created by the concept’s foundations in

the idea of bureaucratic *potential* or *ability* are analytical, not semantic, and these cannot be addressed by simply substituting another catch-all term. It is important to once again emphasise that this paper does not advocate abandoning the concept of bureaucratic capacity entirely, nor are these critiques and suggestions are not aimed at invalidating or replacing the many useful insights generated by research on state capacity. But as this research programme has built increasingly compelling evidence that the quality of the state bureaucracy matters for such outcomes, scholars have rightly sought to answer questions that are increasingly specific to particular bureaucracies, policies and contexts, and it is here that the limitations of capacity as a conceptual framework have become evident.

Moving 'beyond' capacity does not therefore mean abandoning the concept entirely. Rather, it is an argument for greater conceptual precision, and for a broader change in how scholars and practitioners think about policy implementation, bureaucratic performance and reform. While a comprehensive methodological discussion is in itself beyond the scope of any one article, this section sketches three non-mutually exclusive approaches that scholars and practitioners alike can – and in some cases, have already begun to – take.

6.1 Analysing multiple actors and principals

First, research on organisational performance and reform can explicitly engage with the implications of bureaucracies being collective actors under multiple principals. Whereas the approach of state and organisational capacity is to acknowledge these complexities but subsume them into a single concept in search of conceptual simplicity (e.g. Centeno *et al.*, 2017; USAID, 2017), an alternative approach is to centre them in the analysis in order to understand how issues of collective action and collective choice shape organisational performance. A growing number of innovative micro-empirical studies have begun to investigate these issues; pursuing them more deeply and reflecting on their implications for organisational action provides a rich agenda for further research.

While many authors have studied the impact of different information, monitoring, and incentive schemes in the public sector, including through an increasing literature on field experiments (see various in Finan *et al.*, 2017), these are overwhelmingly conducted as problems of a single principal over-seeing a single agent (or multiple agents who do not interact). Given the collaborative nature of many public sector tasks, and the organisation-level evidence that issues of collaboration amongst agents within an organisation are significant for performance (Rasul *et al.*, forthcoming), this is a significant gap – and opportunity for new research. Inspiration here can be drawn from the growing literature on social incentives in organisations (Ashraf and Bandiera, 2017), as well as from classic literature in economics and organisation theory on the complex nature of bureaucratic incentives (e.g. Cyert and March, 1963; Niskanen, 1968).

Similarly, there has been little research on (sub-)optimal allocation of individual talent in the public sector, although a handful of recent studies demonstrate that this is fertile ground. For example, Best *et al.* (2018) and Fenizia (2020) document bureaucrat-level variation in productivity, using samples of public procurement specialists and managers, respectively, and show how optimal allocation of these individuals could lead to significant overall improvements in performance. In another vein, Khan *et al.* (2018) experimentally rotate tax inspectors in Pakistan. While the authors' main goal is to examine whether such the rotation policy elicits additional effort, it illustrates the point that worker–job matching is a powerful determinant of performance. Once again, research on personnel allocation in the public sector can take inspiration from studies on private firms: for example, Bandiera *et al.* (2009) show that social connectedness of supervisors and workers on a farm has a significant impact on worker productivity.

Empirical studies of relational contracts and organisational culture in government organisations are even rarer, although here the scope for positive gains is perhaps the largest. While organisational culture has long been cited as an important factor in organisational performance in developing countries (Grindle, 1997; McDonnell, 2017), its qualitative importance has yet to be matched by quantitative studies of the impact of organisational culture on performance in public sector organisations.

Again, the private-sector side of organisational economics can serve as a source of inspiration, with studies by Martinez *et al.* (2015) and Blader *et al.* (2019) demonstrating the importance of organisational culture for organisational performance.

The challenges imposed on public managers by the multiple principals' competing, unstable and collectively incoherent political demands, and their relationship to efficiency and policy implementation, is another fruitful avenue for the study of bureaucracies in development. These dynamics have also been extensively documented in qualitative literature focused on US public administration (Miller and Whitford, 2016; Pressman and Wildavsky 1973; Wilson, 1989). More recently, quantitative research on policy implementation in developing countries has begun to explore similar themes (Gulzar and Pasquale, 2017; Williams, 2017), with implications for institutional design of accountability systems for bureaucrats and for delivery mechanisms for aid and inter-governmental transfers. At the comparative, historical level, work on polycentric sovereignty by Salter and Young (2018, 2019) documents how polycentric governance arrangements were complementary to the development of effective states in medieval and early modern Europe.

In some cases these multiple principals of public agencies include what Ménard (2014) calls 'meso-institutions' – bodies like central audit authorities, budget offices and public service commissions. Meso-institutions embed specific government agencies into their broader institutional context, and are thus an important component of institutional quality. However, they are often also tangible organisations (as in the case of public sector central management agencies), making them more feasible to actually reform than less tangible social, political and economic institutions. Studying the impact of such meso-institutions on public bureaucracies in terms of their impacts on bureaucratic incentives behaviour, agency governance and balancing of internal and external stakeholders, and performance would help connect micro, meso and macro perspectives on state bureaucracies; analysing their impacts in terms of bureaucratic capacity seems a rather blunt tool in comparison.

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 emphasises 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). Recognising the limitations of capacity for thinking about bureaucratic performance, policy implementation and reform thus opens up avenues for future research and policy innovation that are both practical and theoretically well-grounded.

6.2 Engaging with specificity and contingency

A second way in which scholars can make their analysis better suited to the realities of policy implementation is to engage directly with the complexity and uncertainty of policy implementation, which are not well represented by a single construct like capacity at the national, sub-national or even organisational level. By asking specific questions about the likely outcomes of specific bureaucracies implementing specific policies in specific contexts, scholars and reformers alike can better understand and predict policy implementation and identify specific levers for meaningful improvement. The importance of these contextual specificities and contingencies becomes clear as abstract discussions of state or organisational capacity are narrowed to specific policy questions, as the two cases below illustrate.

First, in research on post-conflict security, agreement implementation and policing – a topic which has spawned a significant literature of cross-country research (Cole, 2015; Englehart, 2009) and case studies (Baker, 2009; Hills, 2007) and which is central to theories of the long-term development of state capacity – a slippage between the macro-historical and development practice conceptions of

capacity is often especially apparent. For example, Hills (2007: 405) states that ‘Police governance is analysed in terms of institutional capacity and technical proficiency’, Baker (2009: 184) defines ‘government capacity’ as ‘the degree of capacity to provide state policing and to regulate, audit, and facilitate other policing agencies’, and Cole (2015) emphasises that states are not unitary actors but goes on to study treaty enforcement as a matter of state capacity. The reductive framing of the political and bureaucratic determinants of government action as matters of capacity is often even more stark from practitioners. For example, Friedman’s (2011: 13) study of post-conflict policing in Liberia quotes a ‘deputy UNMIL police commissioner, [who] said he thought the police possessed “a significant level of capacity and promise”, including a group of mid-level managers with six or seven years of experience that had the skills and integrity to step into the role of inspector general. However, he was critical of the high-level political appointees because such appointments raised the possibility of political meddling and because several appointees had little policing expertise. He said, “It’s a contradiction to try and build police capacity when the top level has no police capacity”’.

For academics and practitioners alike in the study of post-conflict security, then, the slippage of the concept of capacity from its macro-historical origins to its application to specific policies and organisations to individual bureaucrats can mask a far more complex range of processes and outcomes. However, the potential for more nuanced theoretical approaches to provide new insight by disaggregating these analyses is illustrated by Medie’s (2018: 137) study of post-conflict police reform in Liberia and Côte d’Ivoire, which ‘demonstrates that even in the most unlikely of contexts the formal institutions of the state can be made to work in order to deliver better outcomes for marginalised groups’. By analysing the nuances of domestic political coalitions and heterogeneity among government agents in these two contexts that would be generally considered to have relatively low state capacity, Medie is able to not only uncover unexpected policy implementation outcomes but also explain them and their implications for reformers in other contexts.

Second, the case of the failed launch of the [healthcare.gov](https://www.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 due to 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 4 and 5. Although the government had many individuals with appropriate technical skills (and had the resources to hire many others), implementation was characterised 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 organisations. While the US state might have high capacity in a broad, aggregate, cross-national sense, numerous studies of policy implementation in the US emphasise how complexity and contingency can be even more powerful determinants of policy implementation (e.g. Pressman and Wildavsky, 1973).

There are, of course, numerous empirical studies of state bureaucracies that do pay close attention to the idiosyncrasies of their context. To cite but one example, Weaver (2020) uses innovative data on bribery to study corruption in the hiring of public health workers in an anonymised country and finds that in this context corruption actually increases the quality of bureaucratic hires, albeit with substantial variation, and identifies the parameters that determine whether corruption will have a positive or negative impact in a given context. This serves to illustrate the value of asking specific questions about government bureaucracies and answering in ways that engage with contextual details in micro-empirical research. At the same time, it also emphasises the specificity and

contingency of bureaucratic effectiveness, and thus the analytical limitations of approaches that abstract from this.

6.3 Distinguishing between performance and capacity

A final implication of this paper's argument is that theory and measurement should clearly distinguish between performance and capacity when studying state bureaucracies. The two terms (along with others like effectiveness or quality) are often used interchangeably. For instance, the influential Worldwide Governance Indicators' measure of Government Effectiveness is listed as an indicator of the 'capacity of the government to effectively formulate and implement sound policies' (Kaufmann *et al.*, 2010: 4). Research and practice alike would benefit from more clearly and consistently specifying which of these concepts they are aiming to measure and why.

One important distinction is that whereas bureaucratic performance can be measured directly, capacity can only be measured in hypothetical terms. Whereas performance is a *retrospective* measure of bureaucratic quality, capacity is a *prospective* measure. Performance aims to measure what actions or outcomes a bureaucracy actually produced in the past, whereas capacity aims to measure the ability to implement hypothetical policies in the future. The latter is more inherently speculative, as well as dependent on the specific details of the future hypothetical policy. To accurately predict the implementation of a future policy one must not only specify the policy, but also the other contingent circumstances surrounding the implementation. In contrast, measuring performance is free from such hypotheticals, measuring what the bureaucracy actually did than what it might do.

While a full review of the strengths and weaknesses of existing national-level measures of state capacity is beyond the scope of this paper, the excellent surveys by Hendrix (2010), Cingolani (2019) and Hanson and Sigman (2019) demonstrate both the blurred distinction between performance and capacity as well as the challenges of measuring a generic capacity. Regarding the former, Hanson and Sigman (2019: 9) note that 'A common way to measure administrative capacity is to look at the outcomes of public goods and service delivery such as the percentage of children enrolled in primary schools, infant mortality rates, or literacy rates', but point out that this is problematic because a state may simply not prioritise a particular area. There are, of course, some potential measures (such as resource availability) that can reasonably be assumed to reflect a bureaucracy's generic potential capacity because they can be redeployed across a range of policy goals. However, even the oft-used measure of the tax revenue/GDP ratio – perhaps the simplest and most generic measure of potential capacity, at national level at least – has an imperfect relationship to a bureaucracy's hypothetical capacity. While it is doubtless an indicator of the financial resources available to the state, at least in the short-run, not only might different states make policy or ideological choices about tax rates or the optimal size of the state (Hanson and Sigman, 2019: 7), but this ratio reflects not just the actions of notional capacities of the bureaucracy itself but also issues of tax morale and the nature of relationships between state and society.⁷ Indeed, the challenge of measuring capacity objectively is one reason why some of the most widely used measures of state capacity are based on subjective expert opinion surveys, for all their well-recognised limitations (Cingolani, 2019; Hanson and Sigman, 2019). This is not to say that these or other measures of state capacity are orthogonal to either the hypothetical capacity or actual performance of state bureaucracies; rather, the point is that not only do many measures of state capacity blur the distinction between performance and capacity, but also that the interpretation of even the simplest measures of capacity with respect to the core concept is contestable.

Of course, measuring performance in government organisations poses numerous conceptual and practical challenges. Public sector outputs are typically non-priced and are often public goods, which complicates the measurement of quality and value-added (Wilson, 1989). Both individual bureaucrats and organisations as a whole must simultaneously work on multiple tasks (Holmstrom and Milgrom, 1991) for multiple principals (Dixit, 2002), so allocating effort and assessing

⁷I am grateful to an anonymous referee for this point.

performance across different dimensions requires trade-offs and judgement about priorities. Many important bureaucratic tasks require team production (Alchian and Demsetz, 1972; Bandiera *et al.*, 2009), so individual- or even team-level measures of performance may be incomplete or impossible, or require discretionary judgements about the state of the world (Prendergast, 2003) which cannot be externally verified. Bureaucratic action may be difficult to connect to policy outcomes, and differentially so across task and agency types (Wilson, 1989). Even in cases where these conceptual challenges can be surmounted, researchers and managers often encounter the practical obstacle that bureaucracies often do not collect data on many relevant dimensions of performance. Although a recent wave of empirical studies have made significant progress in developing innovative and informative measures of bureaucratic performance at both the individual (Bertrand *et al.*, 2020; Best *et al.*, 2018; Chetty *et al.*, 2014; Fenizia, 2020; Khan *et al.*, 2018) and organisational (Rasul and Rogger, 2018; Rasul *et al.*, forthcoming) levels, the very nature of public sector outputs means that performance measurements will always be contestable.

At the same time, these conceptual challenges apply equally to efforts to measure bureaucratic capacity. Given this, measuring performance at least has the advantage of forcing researchers and managers to grapple explicitly with these challenges and trade-offs rather than abstracting them away or relying on subjective expert opinions to make judgements about the quality of outputs or how to weight different dimensions of capacity. Of course, there are occasions in which researchers or practitioners explicitly seek to make these abstractions, such as in the macro-historical study of long-run state development, and for these purposes such summary indicators may be appropriate. But the concept of capacity cannot serve to abstract away from such conceptual nuances and contextual specificities and at the same time claim to be responsive to them, without becoming so stretched as to lose meaning. The argument advanced in this paper is thus not that researchers should never discuss or seek to measure capacity, but that: (1) they should clearly distinguish between the two concepts, in both theory and measurement; and (2) that the advancing our knowledge about the effectiveness of state bureaucracies is likely to require relatively more focus on performance than on capacity than has been the case to date.

Indeed, one interesting path for further research is precisely about the relationship between performance and capacity: to what extent, and in what ways, does a bureaucracy's current performance predict its future performance, or its ability to perform when new policies are introduced or circumstances change? Evidence from private firms shows that present and future performance tend to be positively (but not perfectly) correlated (Gibbons and Henderson, 2013), but we have little longitudinal evidence from the public sector that speaks to this question beyond isolated case studies. Similarly, we know that some types of effective routines can be redeployed to different tasks as the organisation evolves (Nelson and Winter, 1982) while practices that were beneficial in one state of the world may even inhibit adaptation to changing contexts (Brynjolfsson and Milgrom, 2013), but building further theory and evidence on these dynamics in public sector bureaucracies would represent a fertile research agenda.

In addition to guiding future research, the distinction between performance and capacity is important for the design of reforms, especially when combined with a view of bureaucracies as collective actors. The relationship between individual skill development and improved performance in collective actors such as bureaucracies 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 unformalisable and relational aspects of organisational 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 actualised.

This implies a greater focus on improving team dynamics and organisational 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. While many donor capacity building programmes are designed with an awareness of the importance of organisational and institutional factors in bureaucracies' performance, the most commonly used approach to capacity building in practice has remained individualised skill development through trainings and workshops,

despite the scepticism of these very practitioners about the usefulness of such approaches (USAID, 2017). However, this may be beginning to change. For example, USAID has recently begun to shift the focus of its programme measurement away from measures of capacity and towards performance, writing: ‘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’ (2017: 5).

7. Conclusion

Despite the contributions of state capacity to research on the long-run and comparative development of state bureaucracies, the analytical usefulness of capacity is limited for understanding the performance, implementation and reform of specific bureaucracies or policies. The metaphor of bureaucratic capacity obscures the salient fact that bureaucracies are collective actors operating under the direction of multiple political principals, not unitary actors carrying out well-defined policies. As a result of this multiplicity of agents and principals, understanding failures of service delivery or policy implementation requires analysing these problems of coordination, collective action, and collective choice that theories of capacity subsume into one summary concept.

Building on this critique, this paper has suggested three methodological implications for the analysis of bureaucratic performance and policy implementation in specific contexts. First, analysis should focus explicitly on the problems created by bureaucracies being collective actors under multiple principals, rather than seek to abstract from them. Second, analysts should engage directly with contextual specificities and contingencies. Third, measurement and reform should focus on retrospective performance, not hypothetical capacity. Connecting the macro-historical and cross-country literature on state capacity to studies that do engage with these more precise mechanisms and contextual specificities presents both challenges and opportunities for understanding the development of effective bureaucracies. While a handful of existing studies do seek to bridge these levels of scale (e.g. Ang, 2017; Carpenter, 2001; Miller and Whitford, 2016), there is a need for more such work – especially in developing country contexts.

Research on the development and performance of government bureaucracies has made impressive progress in the three decades since studies 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 translate this general understanding into analysis and reform of specific bureaucracies and their policies will require moving beyond the broad concept of state capacity and disaggregating rather than subsuming the complexities of public bureaucracies. Fortunately, many of the theoretical and methodological building blocks for this approach already exist within political science, organisation theory and organisational economics, and much can be drawn from the more theoretically and empirically nuanced 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 efforts at reform have discovered the hard way. Integrating these insights into theory and empirics on policy implementation and reform – and working with policymakers to test them – represents a rich and potentially transformative research agenda.

Acknowledgements. I am grateful for comments from Yuen Yuen Ang, Michael Bernhard, Derick Brinkerhoff, Jennifer Brinkerhoff, Yanilda María Gonzalez, Tobias Haque, Mai Hassan, Dan Honig, David Jacobstein, Julien Labonne, Brian Levy, Zoe Marks, Peace Medie, Dan Rogger, Bo Rothstein, several anonymous referees and seminar participants at Johns Hopkins University SAIS. Any remaining mistakes are my own.

References

- Acemoglu, D. and J. Robinson (2012), *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*, New York: Crown Business.
- Acemoglu, D., C. Garcí a-Jimeno and J. A. Robinson (2015), ‘State Capacity and Economic Development: A Network Approach’, *American Economic Review*, **105**(8): 2364–2409.

- Acemoglu, D., S. Johnson and J. Robinson (2001), 'The Colonial Origins of Comparative Development: An Empirical Investigation', *American Economic Review*, **91**(5): 1369–1401.
- Alchian, A. A. and H. Demsetz (1972), 'Production, Information Costs, and Economic Organization', *American Economic Review*, **62**(5): 777–795.
- An, Y., M. J. Garvin and R. P. Hall (2017), 'Pathways to Better Project Delivery: The Link between Capacity Factors and Urban Infrastructure Projects in India', *World Development*, **94**: 393–405.
- Andrews, M., L. Pritchett and M. Woolcock (2017), *Building State Capability: Evidence, Analysis, Action*, Oxford: Oxford University Press.
- Ang, Y. Y. (2017), *How China Escaped the Poverty Trap*, Cornell: Cornell University Press.
- Anthopolous, L., C. G. Reddick, I. Giannakidou and N. Mavridis (2016), 'Why e-Government Projects Fail? An Analysis of the Healthcare.gov Website', *Government Information Quarterly*, **33**: 161–173.
- Ashraf, N. and O. Bandiera (2017), 'Social Incentives in Organizations', Mimeo.
- Baker, B. (2009), 'Resource Constraint and Policy in Liberia's Post-Conflict Policing', *Police Practice and Research*, **11**(3): 184–196.
- Bandiera, O., I. Barankay and I. Rasul (2009), 'Social Connections and Incentives in the Workplace: Evidence From Personnel Data', *Econometrica*, **77**(4): 1047–1094.
- Baser, H. and P. Morgan (2008), 'Capacity, Change and Performance', Discussion Paper 59B, ECDPM.
- Bersch, K., S. Praça and M. Taylor (2016), 'State Capacity, Bureaucratic Politicization, and Corruption in the Brazilian State', *Governance*, **30**(1): 105–124.
- Bertelli, A., M. Hassan, D. Honig, D. Rogger and M. J. Williams (2020), 'An Agenda for the Study of Public Administration in Developing Countries', *Governance*, **33**(4): 735–748.
- Bertrand, M., R. Burgess, A. Chawla and G. Xu (2020), 'The Glittering Prizes: Career Incentives and Bureaucrat Performance', *The Review Of Economic Studies*, **87**(2): 626–655.
- Besley, T. and T. Persson (2009), 'The Origins of State Capacity: Property Rights, Taxation, and Politics', *American Economic Review*, **99**(4): 1218–44.
- Besley, T. and T. Persson (2011), *Pillars of Prosperity: The Political Economics of Development Clusters*, Princeton: Princeton University Press.
- Best, M. C., J. Hjort and D. Szakonyi (2018), Individuals and Organizations as Sources of State Effectiveness, December 21.
- Blader, S., C. Gartenberg and A. Prat (2019), 'The Contingent Effect of Management Practices', *Review of Economic Studies*, **87**(2): 721–749.
- Bloom, N., R. Lemos, R. Sadun and J. Van Reenen (2014), 'Does Management Matter in Schools', NBER Working Paper 20667.
- Bockstael, E. (2017), 'Critical Capacity Development: An Action Research Approach in Coastal Brazil', *World Development*, **94**: 336–345.
- Bologna Pavlik, J. and A. T. Young (2020), 'Medieval European Traditions in Representation and State Capacity Today', Mimeo, February.
- Briebea, D. (2018), 'State Capacity and Health Outcomes: Comparing Argentina's and Chile's Reduction of Infant and Maternal Mortality, 1960–2013', *World Development*, **101**: 37–53.
- Brinkerhoff, D. and P. Morgan (2010), 'Capacity and Capacity Development: Coping With Complexity', *Public Administration and Development*, **30**: 2–10.
- Bruszt, L. and N. F. Campos (2019), 'Economic Integration and State Capacity', *Journal of Institutional Economics*, **15**: 449–468.
- Brynjolfsson, E. and P. Milgrom (2013), 'Complementarity in Organizations', in R. Gibbons, and J. Roberts (eds.), *The Handbook of Organizational Economics*, Princeton: Princeton University Press, pp. 11–55.
- Carpenter, D. P. (2001), *The Forging of Bureaucratic Autonomy: Reputations, Networks, and Policy Innovation in Executive Agencies, 1862–1928*, Princeton: Princeton University Press.
- Carrera, L. and P. Dunleavy (2010) 'Why Does Productivity Vary Across NHS Hospital Trusts in England? Untangling How Management Competence and the Use of ICTs Shape Hospitals' Performance', LSE Public Policy Group Mimeo, March.
- Centeno, M., A. Kohli and D. Yashar (eds.) (2017), *States in the Developing World*, Princeton: Princeton University Press.
- Chetty, R., J. N. Friedman and J. E. Rockoff (2014), 'Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates', *American Economic Review*, **104**(9): 2593–2632.
- Chuaire, M. F., C. Scartasini and M. Tommasi (2017), 'State Capacity and the Quality of Policies. Revisiting the Relationship Between Openness and Government Size', *Economics and Politics*, **29**(2): 133–156.
- Cingolani, L. (2013), 'The State of State Capacity: A Review of Concepts, Evidence and Measures', AFD-MGSoG UNU-Merit Working Paper Series on Institutions and Economic Growth.
- Cingolani, L. (2019), 'The Role of State Capacity in Development Studies', *Journal of Development Perspectives*, **2**(1–2): 88–114.
- Cole, W. (2015), 'Mind the Gap: State Capacity and the Implementation of Human Rights Treaties', *International Organization*, **69**(2): 405–441.

- Coleman, J. S. (1982), *The Asymmetric Society*, Syracuse: Syracuse University Press.
- Cyert, R. and J. March (1963), *Behavioral Theory of the Firm*, Englewood Cliffs: Prentice-Hall.
- Department for International Development (2009), 'Political Economy Analysis How To Note', July.
- Dincecco, M. and G. Katz (2014), 'State Capacity and Long-Run Economic Performance', *The Economic Journal*, **126**(590): 189–218.
- Dixit, A. (1996), *The Making of Economic Policy: A Transaction-Cost Politics Perspective*, London: The MIT Press.
- Dixit, A. (2002), 'Incentives and Organizations in the Public Sector: An Interpretative Review', *The Journal of Human Resources*, **37**(4): 696–727.
- Dunleavy, P. (1995), 'Policy Disasters: Explaining the UK's Record', *Public Policy and Administration*, **10**(2): 52–70.
- Englehart, N. (2009), 'State Capacity, State Failure, and Human Rights', *Journal of Peace Research*, **46**(2): 163–180.
- Evans, P. and J. Rauch (1999), 'Bureaucracy and Growth: A Cross-National Analysis of the Effects of "Weberian" State Structures on Economic Growth', *American Sociological Review*, **64**(5): 748–765.
- Fenzia, A. (2020), 'Managers and Productivity in the Public Sector'. Mimeo, 29 June.
- Finan, F., B. Olken and R. Pande (2017), 'The Personnel Economics of the State', in A. V. Banerjee, and E. Duflo (eds.), *Handbook of Economic Field Experiments*, Oxford: North-Holland, pp. 467–514.
- Friedman, J. (2011), 'Building civilian police capacity: Post-conflict Liberia, 2003–2011', *Innovations for Successful Societies*.
- Garicano, L. and L. Rayo (2016), 'Why Organizations Fail: Models and Cases', *Journal of Economic Literature*, **54**(1): 137–192.
- Gibbons, R. and R. Henderson (2013), 'What Do Managers Do?', in R. Gibbons, and J. Roberts (eds.), *The Handbook of Organizational Economics*, Princeton: Princeton University Press, pp. 680–731.
- Gingerich, D. (2013), 'Governance Indicators and the Level of Analysis Problem: Empirical Findings From South America', *British Journal of Political Science*, **43**(3): 505–540.
- Grindle, M. (1997), 'Divergent Cultures? When Public Organizations Perform Well in Developing Countries', *World Development*, **25**(4): 481–495.
- Gulzar, S. and B. Pasquale (2017), 'Politicians, Bureaucrats, and Development: Evidence From India', *American Political Science Review*, **111**(1): 162–183.
- Hanson, J. and R. Sigman (2019), 'Leviathan's Latent Dimensions: Measuring State Capacity for Comparative Political Research', Mimeo, June.
- Harbers, I. (2015), 'Taxation and the Unequal Reach of the State: Mapping State Capacity in Ecuador', *Governance*, **28**(3): 373–391.
- Hendrix, C. (2010), 'Measuring State Capacity: Theoretical and Empirical Implications for the Study of Civil Conflict', *Journal of Peace Research*, **47**(3): 273–285.
- Hills, A. (2007), 'Police Commissioners, Presidents and the Governance of Security', *Journal of Modern African Studies*, **45**(3): 403–423.
- Holmstrom, B. (1979), 'Moral Hazard and Observability', *Bell Journal of Economics*, **10**(1): 74–91.
- Holmstrom, B. and P. Milgrom (1991), 'Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design', *The Journal of Law, Economics, and Organization*, **7**(special issue): 24–52.
- Honig, D. (2018), *Navigation by Judgment: Why and When Top Down Management of Foreign Aid Doesn't Work*, Oxford: Oxford University Press.
- Horton, D., A. Alexaki, S. Bennet-Lartey, K. Noële Brice, D. Campilan, F. Carden, J. De Souza Silva, L. T. Duong, I. Kadar, A. Maestrey Boza, I. Kayes Muniruzzaman, J. Perez, M. Somarriba Chang, R. Vernoooy and J. Watts (2003), 'Evaluating Capacity Development: Experiences from Research and Development Organizations Around the World', ISNAR/CTA, IDRC.
- Johnson, N. and M. Koyama (2017), 'States and Economic Growth: Capacity and Constraints', *Explorations in Economic History*, **64**: 1–20.
- Kaufmann, D., A. Kraay and M. Mastruzzi (2010), 'The Worldwide Governance Indicators Methodology and Analytical Issues', World Bank Policy Research Working Paper 5430.
- Khan, A., A. I. Khwaja and B. Olken (2018), 'Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats Through Performance-Based Postings', NBER Working Paper 24383.
- Leaver, C., R. Lemos and D. Scur (2019), 'Measuring and Explaining Management in Schools: New Approaches Using Public Data', Centre for Economic Policy Research Discussion Paper DP14069, October.
- Lee, M., G. Walter-Drop and J. Wiesel (2014), 'Taking the State (Back) Out? Statehood and the Delivery of Collective Goods', *Governance*, **27**(4): 635–654.
- Leibenstein, H. (1966), 'Allocative Efficiency vs "X-Efficiency"', *American Economic Review*, **56**(3): 392–415.
- Leonard, D. (2010), 'Pockets' of Effective Agencies in Weak Governance States: Where Are They Likely and Why Does It Matter?', *Public Administration and Development*, **30**: 91–101.
- List, C. and P. Pettit (2011), *Group Agency: The Possibility, Design, and Status of Corporate Agents*, Oxford: Oxford University Press.
- Mann, M. (1984), 'The Autonomous Power of the State: Its Origins, Mechanisms and Results', *European Journal of Sociology*, **25**(2): 185–213.

- Martinez, E. A., N. Beaulieu, R. Gibbons, P. Pronovost and T. Wang (2015), 'Organizational Culture and Performance', *American Economic Review: Papers Proceedings*, **105**(5): 331–335.
- McDonnell, E. (2017), 'Patchwork Leviathan: How Pockets of Bureaucratic Governance Flourish Within Institutionally Diverse Developing States', *American Sociological Review*, **82**(3): 476–510.
- McDonnell, E. (2020), *Patchwork Leviathan: Pockets of Bureaucratic Effectiveness in Developing States*, Princeton: Princeton University Press.
- Medie, P. (2018), *Global Norms and Local Action: The Campaigns to End Violence Against Women in Africa*, Oxford: Oxford University Press.
- Ménard, C. (2014), 'Embedding Organizational Arrangements: Towards A General Model', *Journal of Institutional Economics*, **10**(4): 567–589.
- Miller, G. and A. B. Whitford (2016), *Above Politics: Bureaucratic Discretion and Credible Commitment*, Cambridge: Cambridge University Press.
- Mkandawire, T. (2017), 'State Capacity, History, Structure, and Political Contestation in Africa', in M. Centeno, A. Kohli, and D. Yashar (eds.), *States in the Developing World*. Princeton: Princeton University Press, pp. 184–216.
- Moreno, J., L. Noguchi and M. Harder (2017), 'Understanding the Process of Community Capacity Building: A Case Study of Two Programs in Yunnan Province, China', *World Development*, **97**: 122–137.
- Morgan, P. J. (2003), 'One more time: just how should we think about the concept of capacity?' ECDPM Occasional Paper No. 1. European Centre for Development Policy Management: Maastricht, The Netherlands; draft, November.
- Muralidharan, K., P. Niehaus and S. Sukhtankar (2016), 'Building State Capacity: Evidence From Biometric Smartcards in India', *American Economic Review*, **106**(10): 2895–2929.
- Nelson, R. and S. Winter (1982), *An Evolutionary Theory of Economic Change*, London: Belknap Press of Harvard University Press.
- Niskanen, W. (1968), 'The Peculiar Economics of Bureaucracy', *The American Economic Review*, **58**(2): 293–305.
- North, D., J. J. Wallis and B. Weingast (2009), *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*, Cambridge: Cambridge University Press.
- Organization for Economic Co-operation and Development (2006), 'The Challenge of Capacity Development: Working Towards Good Practice', DAC Guidelines and Reference Series.
- Pepinsky, T., J. Pierskalla and A. Sacks (2017), 'Bureaucracy and Service Delivery', *Annual Review of Political Science*, **20**: 249–268.
- Prendergast, C. (2003), 'The Limits of Bureaucratic Efficiency', *Journal of Political Economy*, **111**(5): 929–958.
- Pressman, J. L. and A. Wildavsky (1973), *Implementation*, Berkeley: University of California Press.
- Rasul, I. and D. Rogger (2018), 'Management of Bureaucrats and Public Service Delivery: Evidence From the Nigerian Civil Service', *Economic Journal*, **128**(608): 413–446.
- Rasul, I., D. Rogger and M. J. Williams (forthcoming), 'Management, Organizational Performance, and Task Clarity: Evidence From Ghana's Civil Service', *Journal of Public Administration Research and Theory*.
- Salter, A. W. and A. T. Young (2018), 'A Theory of Self-Enforcing Monetary Constitutions with Reference to the Suffolk System, 1825–1858', *Journal of Economic Behavior and Organization*, **156**: 13–22.
- Salter, A. W. and A. T. Young (2019), 'Polycentric Sovereignty: The Medieval Constitution, Governance Quality, and the Wealth of Nations', *Social Science Quarterly*, **100**(4): 1241–1253.
- Schein, E. H. (1985), *Organizational Culture and Leadership*, San Francisco: Jossey-Bass.
- Shavell, S. (1979), 'Risk Sharing and Incentives in the Principal and Agent Relationship', *Bell Journal of Economics*, **10**(1): 55–73.
- Shepsle, K. (1992), 'Congress is a "They," Not an "It": Legislative Intent as Oxymoron', *International Review of Law and Economics*, **12**(2): 239–256.
- Skocpol, T. (1985), 'Bringing the State Back in: Strategies of Analysis in Current Research', in P. Evans, D. Rueschemeyer and T. Skocpol (eds.), *Bringing the State Back In*, Cambridge: Cambridge University Press, pp. 3–43.
- Soifer, H. (2008), 'State Infrastructural Power: Approaches to Conceptualization and Measurement', *Studies in Comparative International Development*, **43**: 231–251.
- Tendler, J. (1997), *Good Government in the Tropics*, Baltimore: Johns Hopkins University Press.
- Teskey, G. (2012), 'Notes on Capacity Development – with a Focus on Fragile and Conflict-Affected States', World Bank.
- Thompson, L. (2013), 'HealthCare.gov Diagnosis: The Government Broke Every Rule of Project Management', *Forbes Magazine*, December 3, <https://www.forbes.com/sites/lorenthompson/2013/12/03/healthcare-gov-diagnosis-the-government-broke-every-rule-of-project-management/#114dda96b7f7> (accessed September 20, 2017).
- Tilly, C. (ed.). (1975), *The Formation of National States in Western Europe*, Princeton, NJ: Princeton University Press.
- Tuomela, R. (1991), 'We Will Do It: An Analysis of Group Intentions', *Philosophy and Phenomenological Research*, **51**(2): 249–277.
- Ubels, J., N.-A. Acquaye-Baddoo and A. Fowler (eds.). (2010), *Capacity Development in Practice*, London: Earthscan.
- Ubels, J., N.-A. Acquaye-Baddoo and A. Fowler (eds.). (2016), *Capacity Development in Practice*, London: Earthscan.
- United States Agency for International Development (2017), 'Local Capacity Development Suggested Approaches: An Additional Help Document for ADS 201', 13 January.

- Weaver, J. (2020), 'Jobs for Sale: Corruption and Misallocation in Hiring', Mimeo, July.
- Weingast, B. and W. Marshall (1988), 'The Industrial Organization of Congress; or, Why Legislatures, Like Firms, Are Not Organized as Markets', *Journal of Political Economy*, **96**(1): 132–163.
- Whitford, A. (2005), 'The Pursuit of Political Control by Multiple Principals', *Journal of Politics*, **67**(1): 29–49.
- Williams, M. J. (2017), 'The Political Economy of Unfinished Development Projects: Corruption, Clientelism, or Collective Choice?', *American Political Science Review*, **111**(4): 705–723.
- Wilson, J. (1989), *Bureaucracy*, New York: Basic Books.
- World Bank (2016), *Towards a New Approach to Capacity Building in Africa*, Mimeo: World Bank.