An Organizational Theory of State Capacity*

Erik Snowberg
University of Utah
University of British Columbia

Michael M. Ting
Columbia University

March 14, 2022

Abstract

A burgeoning literature recognizes that the efficacy of the state is crucial for economic growth and citizen welfare. However, much of that literature abstracts away from the institutional details underlying state capacity. We develop a theory that provides a working definition of state capacity—the ability to handle administrative problems of varying complexity, such as tax collection—and how it is provided and maintained. We conceive of the state as a knowledge hierarchy, or an information-processing institution that passes problems up a set of organizational layers until a layer with the required expertise solves it. Knowledge hierarchies are costly to establish and operate, and politicians differ in policy preferences and public goods valuations. We embed this structure in a simple political economy framework, where politicians may idle parts of the state depending on electoral prospects, thus reducing output. In conjunction with high partisanship, this gives the state designer incentives to distort the state away from efficient levels of capacity and specialization.

*Preliminary. We thank Dana Foarta and seminar audiences at Carlos III Madrid, CUNEF, Princeton, Florida State, UCSD, Yale, the 2019 Berkeley Formal Theory and Comparative Politics, Nottingham Political Economy, and Utah Orgs and PE Conferences for helpful comments.

1 Introduction

Intuitive notions of state capacity have been shown empirically to be important for determining both political and economic outcomes.[1] Rigorous theoretical underpinnings—including a definition of state capacity which can encompass intuitive notions—have lagged behind empirics. In particular, current state-of-the-art theory equates state capacity with realized tax revenue, does not contemplate changes in state capacity over time, and attributes the dynamics of state capacity to macro-level shocks to preferences—such as wars and depressions—rather than political processes.

We propose a broader definition of state capacity—the problem-solving capacity of the state—and, with tools from organizational economics, use it to address some of the lacuna described above. In our model, problems are solved by a bureaucracy (based on the knowledge hierarchy of Garicano, 2000), and the maintenance of that bureaucracy over time, and hence its structure, is the result of political processes. This allows us to distinguish *state capacity*—the number of problems that a state could solve—from *output* which may be limited by political factors and imperfect state maintenance. We show that there is a non-monotonic relationship between state maintenance and the political competitiveness of a society, with state maintenance being highest in politically competitive societies when there is general agreement over policy, and lowest in politically competitive societies in the presence of disagreement. These patterns of maintenance will lead to political distortions in state design; for example, if the state-designer is from a group that values public goods more than society as a whole, it will create structures that insulate the state from political interference to prevent the opposition from severely reducing output.

Our model sheds light on the capacity of contemporary states—from roughly 1945 on, conflict has been less frequent, and consumed a smaller portion of state revenue than in prior periods. Moreover, even in the late modern period the building of state capacity was not a one-way process—it rose more or less continuously in the Britain, while alternatively building and ebbing in France (Brewer 1988). Further, it ties together notions of state capacity with the organizational structures that are the foundations of state capacity—the bureaucracy. Finally, by introducing a more general notion of state capacity, it allows for

[1] See Johnson and Koyama (2017) for a review.

the possibility of conceptualizing states as something more than just their sheer ability to raise revenue, but rather their ability to implement a census, prosecute crimes, or grapple with the problems created by climate change or a global pandemic.

Three sets of facts motivate the main components of our work. First, shortfalls between capacity and output are both common and deliberate. In many developing countries, tax authorities struggle to reach potential revenues (Schreiber 2019), and politicians have acted over time to diminish the revenue collecting capacity of the state. In the U.S., the Internal Revenue Service lost about 25% of its enforcement staff and 15% of its real budget between 2010 and 2017. Unsurprisingly, the audit rate for individual taxpayers plummeted from 0.9% to 0.5%. These reductions did not affect taxpayers evenly. High-income filers saw their audit rates fall in part due to the under-resourcing of auditing and investigative units, while very low-income filers who could claim the Earned Income Tax Credit were audited at higher rates. Thus, the Republican-controlled Congress managed to re-orient the IRS's enforcement resources to avoid one type of tax avoidance in favor of another (ProPublica 2018, 2020, US CBO 2020). This is inconsistent with the extant literature that views state capacity as an investment in extractive capabilities that are never voluntarily relinquished (Besley and Persson 2009, Acemoglu, García-Jimeno, and Robinson 2015).

Second, accounts of state capacity focus on more than the ability to raise revenue, and recognize that organizations—in the form of functioning, non-corrupt judicial and executive institutions—are crucial for its realization. For example, the development literature has recently begun to focus on the role of personnel in helping the state achieve its goals (Finan, Olken, and Pande 2017). Historically, the development of a powerful, centralized Prussian state was accompanied by extensive institutional reforms that could bring the human capital of enlightenment-educated bureaucrats to handle increasingly complex problems (Johnson 1975). In the contemporary world, highly developed states are generally associated with large bureaucracies. As an example, the share of European Union national workforces employed in public service provision (possibly through private sector providers) in 2016 ranged from 20% in Romania to 40% in Denmark and Sweden (Thijs, Hammerschmid, and Palaric 2017).

Third, even similar macro shocks to preferences do not necessarily lead to similar state capacity outcomes. For example, Brewer (1988) traces the evolution of British and French state capacity over the 18th century. During this time, both nations fought numerous wars—often against each other—yet this led to a drastic increase in state capacity and output for the British state, and the eventual bankrupting of the French state and ensuing revolution. Why? Brewer traces the difference to two important factors. First, the British were primarily a naval power, and the French primarily a land power. As navies required persistent funding for maintenance and upkeep of expensive capital (ships), state output was needed even during periods where Britain was not at war. Alternatively, armies could be raised and de-mobilized relatively rapidly. This factor, coupled with a second—a relative lack of internal rivalries within Britain—resulted in a professional and centralized tax bureaucracy which could experiment with a wide array of taxation instruments. The French, instead, relied on less efficient practices, such as tax farming, that would both reward local aristocrats and could be revoked in between wars. Although the shocks to preferences were similar, their translation into state capacity and output were affected by fine details that limited, or allowed, policy persistence.

To incorporate these observations, this paper examines state capacity, and the evolution of state output, when politicians control some structural and operational parameters of state institutions. We consider a society composed of two groups, each with homogenous preferences, that are tied to a party that represents them in government. We will use the terms group and party somewhat interchangeably. Importantly, groups, and the parties or politicians that represent them, are internally homogenous, but have heterogeneous preferences over state output, and have the institutional tools to hobble the state.

As a theoretical foundation, we turn to *knowledge hierarchies*, which offer a simple way to model expertise and production in complex organizations (Garicano 2000, Garicano and Rossi-Hansberg 2006). In the basic formulation, an organization such as a firm faces problems with a difficulty level drawn from a known distribution on $\mathbb{R}$. The organization consists of a series of layers, each of which is endowed with the ability to solve a subset of possible problems. All problems "enter" through the first layer, which solves those within its knowledge set and sends the remainder on to the next layer at a cost. Each successive layer behaves in a similar fashion, until no layers are left. The central trade-off in designing knowledge hierarchies is between personnel and operational costs. Since the workers in each layer must be capable of handling that layer's most complex task, more layers increase specialization and decrease personnel costs. But more layers also increase the average number of times that problems are passed up, increasing operational costs.

Our game begins with a founding period, which represents an initial shock that enables the creation of a new state function. In it, one group establishes the knowledge hierarchy. This is the basic structure of the state, which determines both the extent of its possible problem-solving ability, or its capacity, and its personnel costs. In every subsequent period, each group attains political power with exogenous probability. While in power, the incumbent group receives a signal about the likelihood of maintaining political power, and then decides which layers of the knowledge hierarchy to activate or idle for the following period. Activated layers can solve problems but impose operational costs. We use the simplest possible structure in assuming that layers can be idled and re-activated at zero cost. Personnel and operational costs are shared by both groups through a tax that falls evenly on the members of both parties. Importantly, idle layers still impose personnel costs. This reflects the idea that governmental leaders can temporarily subvert state offices more easily than they can fire civil servants or neutralize public sector unions.

In each period, the problem-solving done by the set of active layers translates directly into the realized production of a public good, or output. Parties differ in their valuations of the public good, and also value the public good more highly when they are in power than when they are not. We refer to the latter difference, resulting from differences in policy preferences between the incumbent and opposition groups, as *partisanship* (Kasara and Suryanarayan 2020). Thus, when deciding whether to diminish or enhance state capacity the group in power faces a trade off. Enhancing state capacity will be beneficial if the incumbent is returned to power, or if the incumbent's partisanship is low. Instability of policy, either through frequent transitions of power, or large differences between the parties in their preferred policies will lead the incumbent to reduce output below the state's capacity.

Our first results take a given state hierarchy and derives the strategies for activating and idling layers. In equilibrium, each party opens more layers as its electoral prospects improve. Because of their higher operational costs, the layers associated with the most difficult problems are idled first, and thus are least likely to be active over time. High operational costs, partisanship, and differences in public goods valuations all amplify the consequences of elections and reduce the chances that a given layer will stay active. As mentioned earlier, there is no general relationship between political competitiveness and a layer's average output, as balanced electorates can either maximize or minimize the probability of idling.

The activation and idling strategies shape the incentives of the party that establishes the state hierarchy. Some of the implications are intuitive. Because of training costs, the state designer constructs a knowledge hierarchy that solves the easiest problems and possibly excludes the most difficult ones. High operational costs add layers and increase the state's specialization. Capacity increases when partisanship is low, and a specialized, high-capacity, high-output state emerges when there is high consensus over the value of public goods.

Our main results show that the political process can distort the design of the state in several ways, especially when the founding group's opposition has low public goods valuation. This breaks the link between capacity and specialization—most notably, the founding group may create a large amount of state capacity, but reduce the number of layers in order to limit opportunities for political interference and keep output high. To understand why, suppose that the opposition dislikes public goods, but is willing to activate the lowest layer of the state. When the founding group's partisanship is low, it reallocates capacity down to the first layer, in order to take advantage of the opposition's willingness to allow the most basic governmental functions. A less specialized, "bottom heavy" state that insulates basic functions results. By contrast, when the founding group's partisanship is high, the founding party shifts capacity away from the first layer and increases specialization instead. This produces a smaller but "top heavy" state that obstructs the opposition from accomplishing too much.

Partisanship also plays a role in the relationship between capacity and the probability of holding power. Holding power is especially important to the state-builder when partisanship is high, and thus capacity is increasing in the probability they maintain power. This relationship becomes weaker, and may even reverse, as partisanship declines. The inverse relation between capacity and the founding party's prospects for staying in power occurs as it shifts capacity to lower layers that are more likely to be activated by the opposition.

Finally, we compare the state realized through political processes to one designed by a social planner. The social planner would never idle a layer, and thus creates a state with none of the preceding political distortions, and output equals capacity. This shows that efficient states tend to be sub-optimally small, and larger states tend to be designed to be less efficient due to political incentives.

This paper complements an extensive literature on state capacity in economics and historical political economy (Tilly 1990, Besley and Persson 2011, Johnson and Koyama 2017, Berwick and Christia 2018). Much of this work focuses on the development of states across broad swaths of time, and focus on war as a driver of capacity. Moreover, the economic literature has focused on the raising of revenue as a measure of state capacity. Both war and the raising of revenue are clearly of first-order importance. To this literature we add the ability to examine the ebbs and flows of state capacity over relatively short time periods, and in response to political factors. Moreover, we explicitly tie state capacity to the instrument of its achievement, the bureaucracy.

The more limited theoretical literature on state and bureaucratic capacity has tended to focus on the state as the product of investment in extractive ability. For example, in Besley and Persson (2009), politicians invest in legal and taxation capabilities in anticipation of future needs, while in Acemoglu, García-Jimeno, and Robinson (2015), central and local politicians jointly invest in extractive capabilities in an environment with local spillovers. Gennaioli and Voth (2015) treat centralization as a main determinant of extraction, and model a ruler's decision over whether to overcome local resistance to centralization. Another approach considers capacity as a parameter in agency problems involving the bureaucracy (e.g., Huber and McCarty 2004, Ting 2011, Foarta 2021). To our knowledge, the existing theoretical work does not explore the implications of internal organizational structure, or the existence of political actors expressly interested in diminishing state capacity.

Our conception of state capacity as problem-solving ability helps to unify a range of related policies or institutional features. In addition to revenue extraction, studies have emphasized legal protections, fiscal centralization, personnel levels, personnel quality, and general administrative resources (Derthick 1990, Besley and Persson 2009, Brown, Earle, and Gehlbach 2009, Grindle 2012, Dal Bó, Finan, and Rossi 2013, Fukuyama 2013, Dincecco and Katz 2014, Bolton, Potter, and Thrower 2016, Garfias 2018). By focusing on the role of organizational structure, we also allow the possibility of integrating the insights of studies that focus on the institutional details of tax collection (Almunia and Lopez-Rodriguez 2018, Bachas, Jaef, and Jensen 2019, Cullen, Turner, and Washington 2021).

Our paper is also complimentary to the large literature on the bureaucracy in political science, and to a more limited extent, economics. In the organizational economics literature, knowledge hierarchies explicitly sidestep the agency problems that are endemic to firms (Garicano 2000). Thus, our paper sidesteps the agency problems that have received frequent attention in the literature on bureaucracy and organization (e.g., Moe 1989, Gailmard and Patty 2012, Gibbons and Roberts 2014). While this naturally limits the range of organizational problems that can be examined, the framework has the virtues of endogenizing internal organizational structure and allowing for a tractable characterization of output. In fact, its central tension between specialization and generalization is reflected in historical decisions about the design of government personnel systems (e.g., Silberman 1993). Our paper also makes a limited contribution to the organizational economics literature as well: aside from Garicano and Rossi-Hansberg (2012), ours is the only one we are aware of to examine knowledge hierarchies across time. While their model is built on a competitive equilibrium framework, ours is built on one with exogenous political frictions.

We organize the paper as follows. The next section introduces the basic model. Section 3 takes the state structure as given and works through how parties activate and idle layers of the knowledge hierarchy. Section 4 uses these results to derive the design of the knowledge hierarchy. Next, Section 5 then examines comparative statics on state capacity, and Section 6 illustrates counterfactuals. Section 7 concludes with a discussion of how our model can be used to understand some of the examples provided above.

2 Problems and Knowledge Hierarchies

We develop a simple model of the establishment and evolution of state capacity over an infinite horizon. State capacity is useful for addressing a range of policy problems depending on how it is deployed by the government. A particularly salient class of policy problems is the collection of tax revenue, which we use throughout as a running example.

The players are two infinitely-lived parties $k \in \{1,2\}$ that represent different groups in society, each consisting of a share r_k of a unit measure of population. Throughout we will use the terms party and group interchangeably. While we refer to these different groups as parties, note that this does not necessarily imply that the society is a democracy: groups may lose or come to power through different means, including, but not limited to, elections. In each period t, party k is in power with probability r_k. Nature begins each period by revealing a realized probability that k will be in power the next period $\rho_{k,t} \in [r_k - \varepsilon, r_k + \varepsilon]$, where $\varepsilon \in (0, \min[r_k, 1 - r_k]]$, $\mathbb{E}[\rho_{k,t}] = r_k$, and $\rho_{k,t}$ is i.i.d. across periods according to distribution $F_k(\cdot)$, which we assume to be uniform. The complementary probability that k's opposition is elected is $\rho_{-k,t} \in [1 - r_k - \varepsilon, 1 - r_k + \varepsilon]$. Both parties discount the payoff in future periods by a factor $\delta \in (0,1)$.

To collect revenue in each period, the state has to solve a continuum of revenue collection problem. Each problem has a difficulty level $z \in Z \equiv [0,1]$, with lower values of z representing simpler problems. Problems are distributed according to a distribution $G(z)$ with corresponding density $g(z)$ on Z. Throughout we assume $G(z) \sim U[0,1]$, but maintain the $G(\cdot)$ notation to clarify the role of the distribution of problems.

Revenue collection is performed by an organization, called a knowledge hierarchy, that divides the bureaucrats that work for it into an integer number $J \geq 0$ of (possibly overlapping or disjoint) layers. Each layer j is associated with a set $Z_j = [z_j, \overline{z}_j] \subseteq Z$ of problems that it can solve. The measure of workers required to address layer j's problems is $G(\overline{z}_j) - G(z_j)$. We will refer to J as the knowledge hierarchy's *depth*, and to Z_j as layer j's *knowledge set*. Knowledge sets are ordered in increasing sequence according to z_j.

A central feature of the model is the endogeneity of the state's effective depth over time. The depth J is established at the beginning of the game, but upon observing its probability of staying in power in the next period, each incumbent may choose which layers will function in the following period. A layer that is *active* can solve problems in its knowledge set, while a layer that is *idle* cannot. Activating or idling layers is costless. Each solved problem results in an equal amount of tax revenue, which can be spent

A knowledge hierarchy solves problems by passing them up through the layers until it is determined that the problem cannot be solved, or it is passed to a layer that can solve the problem. To give a concrete example, suppose that all layers are active in a given period. When a new problem is drawn with complexity z, it is initially processed by layer 1 of the knowledge hierarchy. If $z < z_1$ or $z > \max\{\overline{z}_j\}$, then the problem is dropped from consideration. If $z \in Z_1$, then the problem is solved by layer 1 that period. Otherwise, if $z > \overline{z}_1$, then layer 1 passes the problem on to layer 2. Layer 2 drops the problem if $z < z_2$, addresses the problem if it is in Z_2, or else passes it on to layer 3, and so on until a problem is either solved if z belongs to some knowledge set, or dropped if it does not. Observe that if $z_{j+1} < \overline{z}_j$, then the bureaucrats in level $j+1$ will never process any problems with $z < \overline{z}_j$. Thus, we assume without loss of generality that $z_{j+1} \geq \overline{z}_j$. Let $\mathcal{Z} = \cup_i Z_i$ represent the union of knowledge sets, which is equivalently the set of problems solvable by the knowledge hierarchy. Bureaucrats within a layer cannot discriminate between problems. Problems that do not belong to any knowledge set are ignored, and no problems are solved when $J = 0$.

We define *capacity* as z_J, the hardest problem that the state can possibly solve. In any period, *output* is the measure of all problems that the active layers of the state can actually solve, which may be less than capacity.

The expected utility of a member of party k from solving problems in layer j in period t is:

$$\int_{z_j}^{\overline{z}_j} w_{k,t}^l dG(z) = w_{k,t}^l \left(G(\overline{z}_j) - G(z_j)\right). \tag{1}$$

This benefit depends on the parameter $w_{k,t}^l$, which is the marginal value of government spending for party k in period t when party l is in power. Its possible values are:

$$w_{1,t}^l = \begin{cases} 1 & \text{if } l = 1 \text{ is in power} \\ 1 - \pi_1 & \text{if } l = 2 \text{ is in power} \end{cases} \qquad w_{2,t}^l = \begin{cases} q(1-\pi_2) & \text{if } l = 1 \text{ is in power} \\ q & \text{if } l = 2 \text{ is in power} \end{cases} \tag{2}$$

The parameter $q > 0$ represents the fact that different groups may differ in their valuation of public goods or government policy. As it is possible for $q > 1$, either party could value policy more highly. Thus, it is without loss of generality to consider the case where Party 1 constructs the knowledge hierarchy, as we do in Section 4.

The parameters $\pi_k \in [0, 1]$ represent *partisanship*, or differences in preferences (or ideologies) between the two groups. which we allow to be asymmetric (e.g., Mann and Ornstein 2012, Grossman and Hopkins 2016). If both values are high, this corresponds to an environment in which parties are able to target resources to their supporters, for example when groups are highly aligned with geographic or cultural cleavages. If both values are low, this corresponds to a situation where both parties favor public goods that are (somewhat) valuable to both groups. Finally, if one value is low, and the other is high, this corresponds to an environment where one group favors universalistic programs, while the other prefers to target its own supporters. We say that the groups are *polarized* if some π_k is high.

The state structure imposes two kinds of costs, which are born by the members of both groups, whether or not they are in power. First, a problem of difficulty z must be handled by a bureaucrat with labor cost cz each period, where $c > 0$. As bureaucrats within a layer cannot discriminate among received tasks, they must be able to solve any problem in their knowledge set Z_j, as well as all easier problems. To reflect the difficulty of firing government personnel (whether due to unions, political opposition, civil service rules, or bureaucratic inertia), this cost is fixed when the state is established and incurred for each layer in every period, regardless of whether the layer is idle.[2] The total personnel cost of layer j over all of time is thus

$$\sum_{t=1}^{\infty} \delta^t c\overline{z}_j \int_{z_j}^{\overline{z}_j} dG(z) = \frac{\delta}{1-\delta} c\overline{z}_j \left(G(\overline{z}_j) - G(z_j)\right) \tag{3}$$

Specialization (in the sense of having multiple layers) therefore economizes on total training costs, as fewer workers need to be trained to solve the most difficult problems.

Second, processing each problem imposes operational costs. We implement this in a simple way by assuming a cost $h > 0$ whenever an active layer receives a problem to process, regardless whether it is actually able to solve it. There is no cost for a layer to "send" a problem on to the next layer. Thus, any problem solved by layer j ultimately incurs a search cost of jh. Note that if $jh > w_k^l$, the marginal value of layer j for party k may be negative. The design of the knowledge hierarchy balances the operation and training

[2] This is qualitatively equivalent to c representing a one-time training cost. If labor costs are not incurred when a layer is idled this will change the incentives somewhat for idling and activating layers, although a suitably modified version of Proposition 1 will still hold. Other propositions will be qualitatively unaffected.

costs: many layers saves on labor costs, but creates higher operational costs, whereas few layers saves on operational costs, but at the expense of higher labor costs.

2.1 Timing

The party power in period $t = 0$ can create the structure of the state $\mathcal{Z}$ that will operate for the remainder of the game. The strategy space for Party 1 in period $t = 0$ is the powerset of all compact intervals of $[0, 1]$. The timing of the rest of that stage, and the ones that follow is:

1. $\rho_{k,t}$ is realized.

2. The party in power determines which layers of the state will be idled or active in the next period.

3. A power transition may occur, with the party in power determined according to probability $\rho_{k,t}$.

4. The party in power in the next period raises revenue with the revenue collection agency that they inherit, and allocates that revenue towards their policy aim(s).

The party k stationary strategy for activation and idling in each period is denoted $\mathbf{a}_k : [0, 1] \to \{0, 1\}^J$.

3 State Maintenance

As our opening discussion illustrates, existing state structures are subject to political manipulation. We therefore begin by characterizing the stationary strategies for activating and idling the layers of a given knowledge hierarchy $\mathcal{Z}$.

3.1 Activating and Idling

Because idling and activating are costless, the current set of layers in either state is irrelevant for future payoffs. As such, we derive a strategy that only depends on the characteristics of the knowledge hierarchy (number of layers, and their boundaries), and the probability that party k is in power in the following period(s) $\rho_{k,t}$. The next section uses these results to solve for the optimal knowledge hierarchy designed at $t = 0$, given these strategies.

The period $t+1$ expected value of an active jth layer $Z_j = [z_j, \overline{z}_j]$ to an incumbent party k is the measure of problems solved by that layer $(G(\overline{z}_j) - G(z_j))$, times the value to individuals in that group—w_k^l, which will depend on who is in power in the next period—of public goods paid for with the revenue generated from solving that measure of problems:

$$(\rho_{k,t} w_k^k + (1-\rho_{k,t}) w_k^{-k} - jh)(G(\overline{z}_j) - G(z_j)).$$

As personnel costs (c) are effectively sunk, the marginal value of problems solved in a given layer depend only on communication costs, which increase with a layer's depth j. As such, only the highest layers of any knowledge hierarchy will be idled.

Define p_k^j as the realized value for party k of $\rho_{k,t}$ that makes it indifferent between keeping either the lowest $j-1$ or lowest j layers active. Denote by $\overline{v}_k$ and $\underline{v}_k$ the ex ante average payoff party k expects conditional upon winning and losing, respectively. Finally let $\overline{v}_k^j$ and $\underline{v}_k^j$ denote the continuation value conditional upon winning and losing, respectively, when the party in power in the current period inherits j active layers.

This produces a system of $2(2J-1)$ linear equations. The expected value for each j and k can be written as[3]

$$\begin{aligned}
\overline{v}_k^j &= \sum_{i=1}^{j} (w_k^k - ih)(G(\overline{z}_i) - G(z_i)) + \delta(r_k \overline{v}_k + (1-r_k)\underline{v}_k) \\
\underline{v}_k^j &= \sum_{i=1}^{j} (w_k^{-k} - ih)(G(\overline{z}_i) - G(z_i)) + \delta(r_k \overline{v}_k + (1-r_k)\underline{v}_k).
\end{aligned} \tag{4}$$

Next, the ex ante expected values are:

$$\begin{aligned}
\overline{v}_k &= \sum_{j=1}^{J-1} \left(F_k(p_k^{j+1}) - F_k(p_k^j) \right) \overline{v}_k^j \\
\underline{v}_k &= \sum_{j=1}^{J-1} \left(F_{-k}(p_{-k}^{j+1}) - F_{-k}(p_{-k}^j) \right) \underline{v}_k^j .
\end{aligned} \tag{5}$$

[3]Note that both expressions omit personnel costs c, as those are incurred whether a party wins or losses, and hence are not important to the analysis of idling and activating.

Under the assumption that realized election probabilities are distributed uniformly on $[0,1]$, the parenthesized expressions in (5) become $p_k^{j+1} - p_k^j$ and $p_{-k}^{j+1} - p_{-k}^j$, respectively.

For any j between 1 and J, p_k^j—the probability of returning to office that makes k indifferent between keeping $j-1$ and j layers active—must satisfy:

$$p_k^j \overline{v}_k^{j-1} + (1-p_k^j)\underline{v}_k^{j-1} = p_k^j \overline{v}_k^j + (1-p_k^j)\underline{v}_k^j, \tag{6}$$

Note that as p_k^j is a stationary indifference condition, it is independent of the distribution of election probabilities. As such, we can then define the expected probability—before ρ_k is realized—party k activates layer j as $\sigma_k(j) \equiv 1 - F_k(p_k^j)$. Solving for p_k^j produces the following interior probability cutoff:

$$p_k^j = \frac{w_k^{-k} - jh}{w_k^{-k} - w_k^k} = \begin{cases} -\dfrac{(1-\pi_1) - jh}{\pi_1} & \text{for Party } k = 1, \\ -\dfrac{q(1-\pi_2) - jh}{q\pi_2} & \text{for Party } k = 2, \end{cases} \tag{7}$$

where (7) follows from substituting in the value functions (4) into (6).

Expression (7) and the associated probabilities of activation $\sigma_k(j)$ are essential for understanding both state maintenance in this section, and state design decisions in Section 4. First, these thresholds, and the associated probabilities of activation, are fixed across periods, as expected in a stationary equilibria. Second, they are independent of the distribution of problems $G(z)$. Perhaps more importantly, the thresholds for a given party vary only in response to j—with higher j's requiring a higher threshold, and thus having a lower probability of opening. A third useful fact immediately apparent from (7) is that if $jh > 1 - \pi_1$, then Party 1 will never activate layer j along the equilibrium path. As Party 1 will not design a state to include a layer it never uses, we can use this fact to establish a loose, but useful bound on J: $J < (1-\pi_1)/h$ (and hence $jh < 1 - \pi_1$ for all $j < J$).

Note further that, for both parties, the denominator of (7) is always negative. Thus, if $w_k^{-k} > jh$, party k will always activate layer j. If, on the other hand, $w_k^{-k} < jh$, then the cutoff will be increasing, and probability of activation decreasing, in partisanship π_k.[4]

[4]This is true if $1 - jh > 0$, which is implied by $J < 1/h$.

Intuitively, partisanship makes state maintenance less attractive as spending when the other party comes to power will be outweighed, to an even greater extent, by the cost of the state maintenance. Finally, note that the largest difference between the two parties in their thresholds comes from the asymmetry of valuation of public goods q. The thresholds of Party 2 are decreasing—and corresponding probabilities of activation $\sigma_2(j)$ increasing—in q. So when partisanship is symmetric ($\pi_1 = \pi_2$), and Party 2 values public goods less than Party 1 ($q < 1$), then Party 2 will be less inclined to maintain the state. On the other hand, when $q > 1$, Party 2 will be more inclined to maintain the state. We return to formalize these intuitions, in Section 3.3, after examining an example of how activating and idling affects value functions in the next subsection.

3.2 Example of Value Function

The overall value of the activation/idling decision is convex, although the value of activating or idling any one layer is linear. We illustrate why with a simple example, in Figure 1. This figure plots the $\overline{v}_k^j$ value functions as a function of realized probability of retaining power ($\rho_{k,t}$) for a three-layer knowledge hierarchy with contiguous layers $[0, 0.2)$, $[0.2, 0.3)$, and $[0.3, 0.35]$. Election probabilities are symmetric with $\rho_{k,t} \sim U[0, 1]$. This figure sets $\pi_1 = \pi_2 = 0.1$ and $q = 1$, and thus payoffs and value functions for both parties are identical.

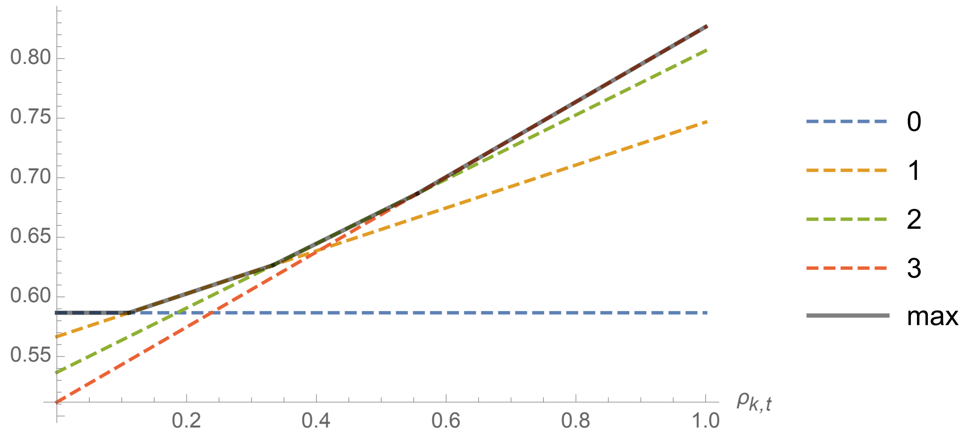


Figure 1: Value Functions by Depth, $J = 3$. Each line plots continuation value for 0, 1, 2, 3 layers, given equilibrium future choices. Parameter values are $h = 0.2$, $r_k = 0.5$ ($\rho_{k,t} \sim U[0,1]$), $q = 1.0$, $\pi_1 = \pi_2 = 0.1$, and the knowledge hierarchy layers are $(0, 0.2, 0.3, 0.35)$. The p_k^j cutoffs for choosing 1, 2, and 3 layers are 1/9, 3/9, and 5/9, respectively.

As previously discussed, the optimal way to maintain j layers in a depth-J knowledge hierarchy is to activate the j lowest layers and idle the $J - j$ highest ones. The figure shows that for $j > 0$ each $\overline{v}_k^j$ is increasing in $\rho_{k,t}$: the value of an active layer is higher when the politician expects re-election. Additionally, the slope of $\overline{v}_k^j$ is increasing in j, which reflects the fact that payoffs are more volatile when more layers are active. Thus the optimal number of active layers increases in $\rho_{k,t}$. Upon learning $\rho_{k,t}$, the incumbent party simply selects the value of j that maximizes $\overline{v}_k^j$, as traced out by the black line in the figure.

3.3 Comparative Statics on Maintenance

We now examine how average activation probabilities change with respect to various parameters. These probabilities will affect the period 0 value of constructing a layer of given depth and width—in addition to some parameters affecting this decision directly—as discussed in the next two sections.

Along the equilibrium path, layer j's expected output is its probability of activation, or:

$$\sum_k r_k \sigma_k(j) = 1 - r_1 F_1(p_1^j) - r_2 F_2(p_2^j). \tag{8}$$

That is, the average probability that a layer is active is simply the probability that Group

1 has a realized probability of maintaining power high enough to warrant (re-)activation, multiplied by the probability that the group is in power in the first place, plus the probability that Group 2 has a realized probability of maintaining power high enough to (re-)activate that layer, times the probability that they are in power.

Our first result uses this expression to establish some basic comparative statics on the probability of activating some layer j for a given state structure.

Proposition 1 [Activating and Idling.] *For a state represented by knowledge hierarchy $\mathcal{Z}$, the probability that layer j is active is:*

(i) Weakly decreasing in h and π_k, and weakly increasing in q,

(ii) 0 for an interval of r_1 only if $q < \frac{jh\pi_1}{\pi_1+\pi_2((1-jh)-\pi_1(1-2\varepsilon))}$. This interval is internal to (0,1) only if $jh < q(1+\varepsilon\pi_2)$.

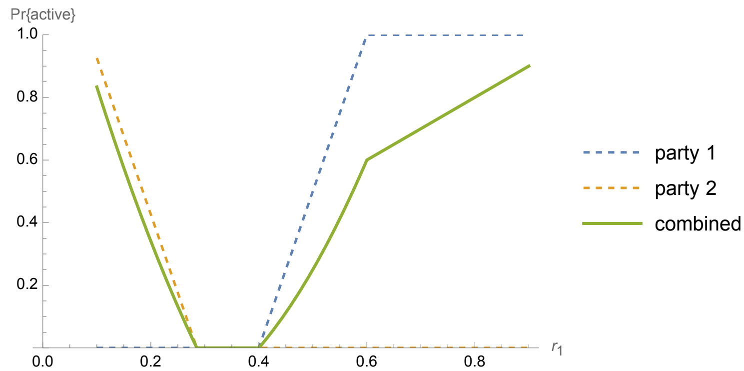
(iii) 1 for an interval of r_1 only if $q > \frac{jh\pi_1}{\pi_1+\pi_2((1-jh)-\pi_1(1+2\varepsilon))}$. This interval is internal to (0,1) only if $jh > \max[1-\pi_1(1+\varepsilon), q(1-\pi_2(1+\varepsilon))]$.

A number of parameters have straight-forward and intuitive effects on state maintenance, as described in part (i) of the proposition. With high partisanship (π_k), and low public goods valuations (q), the prospect of an opposition-run state can be unappealing an incumbent parties. An incumbent that will likely lose power will thus be inclined to idle the highest layers of the state. High operational costs (h) reduce the value of active layers further and therefore have an effect similar to increasing partisanship.

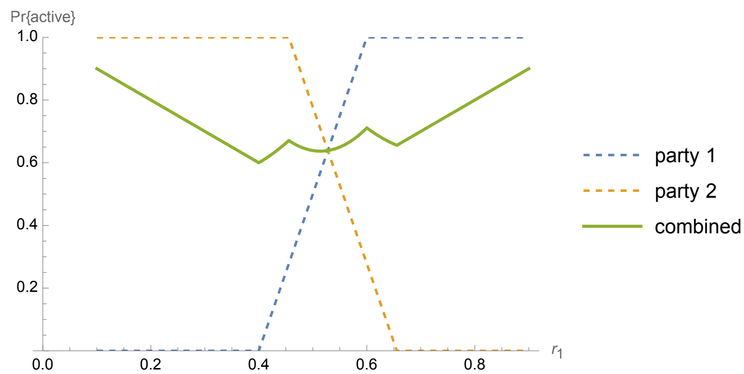
On the other hand, state maintenance is always non-monotonic in the probability that one party or the other (r_k) remains in power, although moderate values of those parameters may either minimize or maximize realized output. In particular, parts (ii) and (iii) show that moderate values of r_k can be associated with a layer being always idled ($\sigma_k(j) = 0$) or always activated ($\sigma_k(j) = 1$). These occur when $r_k + \varepsilon \leq p_k^j$ or $r_k - \varepsilon \geq p_k^j$, respectively. Parts (ii) and (iii) then follow from finding conditions under which the ranges of r_k overlap, but do not contain the entire $(0, 1)$ interval.

To understand why moderate values of r_k can be associated with a layer being either always idled or always activated, we focus on the public-goods valuation (q) of Party 2, and illustrate its workings in Figure 2. When Party 2 does not value public goods highly (q low)

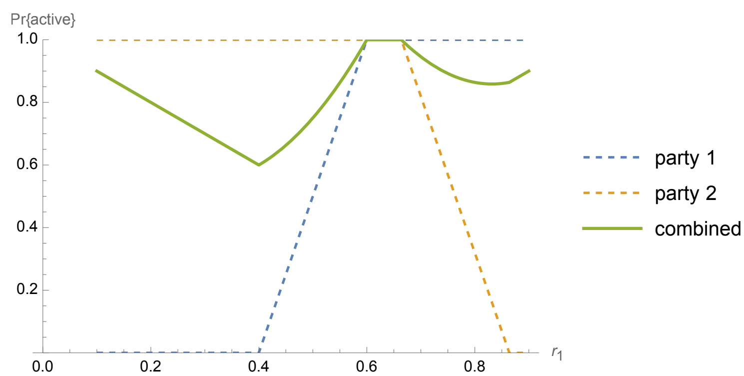
Figure 2: Maintenance may be highest or lowest in societies with a high-probability of power transitions, depending on q.



Panel A: Low q



Panel B: Moderate q



Panel C: High q

Notes: All three panels use $j = 1$, $h = 0.5$, $\varepsilon = 0.1$, $\pi_1 = 1$, and $\pi_2 = 0.9$. In Panel A $q = 0.6$, in Panel B $q = 1$, and in Panel C $q = 1.6$.

then it will only activate layers when it is relatively sure of re-election. At the same time, if π_1 is high, Party 1 will only activate the layer if it is relatively sure to stay in power. This implies that a given level of the knowledge hierarchy may cease function when r_1 is moderate, as shown in Panel A. In other words, a dominant party or group is needed to maintain a highly-functional state. On the other hand, when Party 2 values public goods at a relatively equal level to Party 1, it will become willing to keep all layers active, even when its chances of staying in power are modest, making moderate values of r_1 best for state maintenance, as shown in Panel C. In between these two extremes, moderate probabilities of maintaining power lead to moderate levels of state maintenance, as shown in Panel B.

Election probabilities also interact with the other parameters. High communication costs (h) and high partisanship (π_k) make parties less willing to keep layers activated unless they are relatively sure of holding onto power, creating scenarios where a dominant party or group is needed for state maintenance, as in Panel A. On the other hand, if communication costs are low, and/or the public goods valuations are high, then maintenance will be highest when both parties have a moderate chance of being in power, as in Panel C.

The election-induced incentives to maintain the state have clear implications for realized output and hence state design. The following remark shows that Party 1's partisanship can drive a wedge between capacity and output in any state that it would create, as it would dislike layers that Party 2 would activate. The bound follows simply from manipulating $r_1\sigma_1(j)(1-jh) + r_2\sigma_2(j)(1-\pi_1-jh)$, or Party 1's ex ante expected value of problems solved in layer J, and thus we state the result without proof.

Remark 1 [Output and Capacity.] *If $\pi_1 > 1 - jh$, then for any layer j designed by Party 1:*

$$\sigma_2(j) < -\frac{r_1\sigma_1(j)(1-jh)}{r_2(1-\pi_1-jh)}.$$

Party 1 takes these relationships into account when designing the state. In particular, when its partisanship is high, it may shrink layers that Party 2 will activate and expand those that it will idle. Likewise, lower partisanship will give it an incentive to increase the size of layers that Party 2 is likely to activate. The subsequent sections examine the implications of these incentives for the capacity and specialization of the state.

4 State Design

We proceed in two steps to solve for the optimal knowledge hierarchy / knowledge sets $\mathcal{Z}$, given the opening and closing strategies described above. First, Section 4.1 establishes some simple results useful for both characterizing the optimal hierarchy, and for providing an intuition for its structure. In particular, the optimal knowledge hierarchy will have no "gaps" (that is $\overline{z}_j = z_{j+1}$), that the lowest knowledge set will contain 0 (that is, $z_1 = 0$), and that knowledge sets will be decreasing in length. Thus, to build the optimal knowledge hierarchy one can anchor it at zero, and then stack layers on top of it until an optimal next layer would have negative length (that is, $\overline{z}_{J+1} < \overline{z}_J$). At that point, J, the last layer that the state designer wishes to assign positive length, is the top layer of the optimal hierarchy. Second, Section 4.2 uses these results to explicitly define the optimal structure.

4.1 Basic Results

The initial design of the knowledge hierarchy depends Party 1's anticipation of which layers will stay active over time. As established above, parties determine which layers to activate or idle for the following period based on their realized probability of holding onto power. Only the lowest layers are activated, and the reelection probability cutoff p_k^j for activating layer j is given by (7).

We next establish the optimal state structure from the perspective of the Party 1 in the initial period ($t = 0$). Our first result greatly simplifies the analysis by pinning down the possible types of solutions. In an optimal knowledge hierarchy, knowledge sets are "stacked," or arranged sequentially with no overlaps and no gaps in between. These stacks are "anchored" at 0 to emphasize the easiest (low z) problems. (The proof of this and all other results can be found in the Appendix.)

Lemma 1 [Stacking and Anchoring Knowledge Sets.] *In an optimal depth-J knowledge hierarchy,* $\overline{z}_j = z_{j+1}$ *for* $j = 1, \ldots, J-1$ *and* $z_1 = 0$*.*

Lemma 1 simplifies the derivation of optimal knowledge hierarchies by allowing us to restrict attention to stacked and anchored knowledge sets. Accordingly, we adopt the notation that each Z_j takes the form $[\overline{z}_{j-1}, \overline{z}_j]$ for each layer j. These knowledge sets form a partition of $[0, \overline{z}_J]$, where $\overline{z}_J$ is the capacity of the state.

The party in power in period $t = 0$ thus has the following maximization problem:

$$\max_{J,\overline{z}_0,\ldots,\overline{z}_J} U_1(\mathcal{Z}) \quad = \quad \sum_{t=1}^{\infty} \delta^t \sum_{j=1}^{J} \left[\sum_{k=1}^{2} r_k \sigma_k(j)(w_1^k - jh) - c\overline{z}_j \right] (G(\overline{z}_j) - G(\overline{z}_{j-1})) \,, \qquad (9)$$
$$\text{s.t.}: \ \ 0 \le \overline{z}_0 \le \overline{z}_1 \le \ldots \le \overline{z}_J \le 1.$$

Note that in (9), $\sigma_k(\cdot)$ depends only on j.

It will be useful to focus initially on the construction of optimal knowledge hierarchies for a given J. While the objective (9) is concave in each z_j, the constraints generally ensure that it is not possible to have a solution where all $z_j \in (0, 1)$. The next result shows that anchoring at $z_0 = 0$ has the fortunate consequence making the objective function globally concave for all remaining z_j terms.

Lemma 2 Concavity of Depth-J Knowledge Hierarchies. *Fixing J and $z_0 = 0$, the party's objective* (9) *is concave over knowledge sets satisfying $\overline{z}_0 \le \overline{z}_1 \le \cdots \le \overline{z}_J$.*

Due to the concavity of $U_1(\mathcal{Z})$, we can solve for the optimal knowledge hierarchy, up to the constraints in (9) by using the first order conditions. For layers $j = 1, \ldots, J-1$, and making use of the fact that $G(z) \sim U[0, 1]$ these can be written as:

$$\begin{aligned}
\overline{z}_j &= \frac{\overline{z}_{j-1} + \overline{z}_{j+1}}{2} + \frac{1}{2c} \sum_{k=1}^{2} r_k \left[\sigma_k(j)(w_1^k - jh) - \sigma_k(j+1)(w_1^k - (j+1)h) \right] \\
&\vdots \\
\overline{z}_J &= \frac{\overline{z}_{J-1}}{2} + \frac{1}{2c} \sum_{k=1}^{2} r_k \sigma_k(J)(w_1^k - Jh).
\end{aligned} \qquad (10)$$

The quantity $(w_1^k - jh)$ is simply Party 1's marginal value of an active layer j when party k is in power. This must be sufficiently positive, in expectation, for Party 1 to create a jth layer.

The expressions in (10) convey a simple intuition for the effect of costs and election probabilities. Since the probability $\sigma_k(j)$ of activating layers does not depend on personnel costs (c), increasing these costs will shift the layer j bound $\overline{z}_j$ downwards toward the midpoint between $\overline{z}_{j-1}$ and $\overline{z}_{j+1}$. Increasing communication costs (h) and partisanship (π_k) have the opposite effect, as these increase threat of idling the $j+1$-th layer and thus give the institution designer an incentive to increase the size of the jth layer.

Solving the system (10) in terms of $\overline{z}_J$ produces a unique interior layer boundary for each knowledge set:

$$\overline{z}_j \quad = \quad \frac{j}{J}\overline{z}_J + \frac{1}{cJ}\sum_{k=1}^{2} r_k \left(J \sum_{i=1}^{j} \sigma_k(i)(w_1^k - ih) - j \sum_{i=1}^{J} \sigma_k(i)(w_1^k - ih) \right). \qquad (11)$$

We can further solve for $\overline{z}_J$, producing the following expression for capacity, which will be correct when $\overline{z}_J \leq 1$:

$$\overline{z}_J = \frac{1}{c(J+1)} \sum_{k=1}^{2} r_k \sum_{i=1}^{J} \sigma_k(i)(w_1^k - ih). \qquad (12)$$

Thus far, we have not addressed the possibility of corner solutions, where $z_j = z_{j+1}$. Optimal knowledge hierarchies with degenerate layers are possible, in part because the state designer may want to prevent (through high operational costs) her opponent from activating layers when she is out of power. Such designs are both relatively rare and cumbersome to characterize, so the subsequent analysis focuses on interior solutions to the state designer's problem.

4.2 Full Characterization

To complete the derivation of the optimal knowledge hierarchy, we determine the optimal depth J^*. The main challenge of this exercise is that J^* can change with the parameters of the model. We thus introduce a general property of optimal layer arrangements that is useful for developing an intuition for the maximum possible depth.

Party 1 prefers J over $J-1$ layers when it is willing to allocate layers of positive measure to layer J. Since it could feasibly add a degenerate Jth layer to an optimal depth-$J-1$ knowledge hierarchy without incurring additional personnel or operational costs, adding any nondegenerate layer must be strictly beneficial. To see when the Jth layer has positive

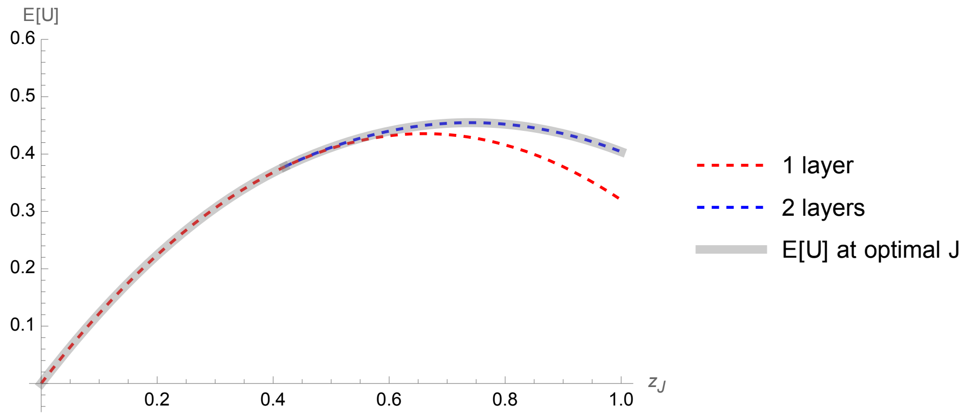


Figure 3: Layers, Capacity, and Expected Utility. Here, $h = 0.2$, $c = 1.0$, $\delta = 0.75$, $\pi_1 = 0.5$, $\pi_2 = 0.75$, $q = 0.25$, $r_1 = 0.5$, and election probabilities are distributed according to $U[0,1]$. The gray curve is the expected utility of a knowledge hierarchy as a function of capacity when J and all internal layers are chosen optimally. Two layers become optimal at $z_J = 0.42$; the optimal knowledge hierarchy is $J = 2$ and $\overline{z}_2 = 0.74$.

size, we solve $\overline{z}_J - \overline{z}_{J-1} \geq 0$ in terms of z_J using expressions (11) and (12). This produces the minimum capacity for a knowledge hierarchy that sustains J layers:

$$\underline{z}_J = \frac{1}{c}\sum_{k=1}^{2} r_k \left(\sum_{i=1}^{J-1} \sigma_k(i)(w_1^k - ih) - (J-1)\sigma_k(J)(w_1^k - Jh) \right). \tag{13}$$

For a state of depth J, a capacity of at least $\underline{z}_J$ is obviously necessary for an interior solution to the state designer's problem.

It is straightforward to show that higher-depth states require larger capacity, and so $\underline{z}_J$ is increasing in J. Once capacity exceeds $\underline{z}_J$, the optimal depth becomes at least J. Thus the optimal depth is J when $\overline{z}_J$ lies in the interval $[\underline{z}_J, \underline{z}_{J+1})$. The weakly monotonic relationship between the optimal depth and problem-solving ability effectively reduces the politician's problem to maximizing over capacity. This fixes the maximum number of non-empty layers for an optimal stacked and anchored knowledge hierarchies, with internal boundaries uniquely characterized by equation (11).

To put this another way, for any capacity level $\overline{z}_J$, there is an optimal number and arrangement of the layers that uniquely minimize costs, given the concavity of the cost function. This arrangement maximizes Party 1's utility, as its benefits are fixed for a given

$\overline{z}_J$. This is illustrated in Figure 3, where one layer minimizes costs for $\overline{z}_J$ less than 0.42 and two layers minimize costs above 0.42. The overall expected utility function, consisting of the curve for $J = 1$ for $\overline{z}_J < 0.42$ and the curve for $J = 2$ for $\overline{z}_J > 0.42$, is smooth and concave.

The final result in our characterization shows the concavity of Party 1's objective even when J is allowed to be chosen optimally, and provides a unique solution that is often straightforward to derive.

Proposition 2 [Optimal Capacity and Depth.] *There is a unique optimal knowledge hierarchy for the Party* 1 *politician. At an interior solution,*

$$\overline{z}_{J^*} = \frac{1}{c(J^*+1)} \sum_{k=1}^{2} r_k \sum_{i=1}^{J^*} \sigma_k(i)(w_1^k - ih), \tag{14}$$

in which J^* is the depth such that $\overline{z}_{J^*} \in [\underline{z}_{J^*}, \underline{z}_{J^*+1})$, or equivalently the value of J such that:

$$\sum_{k=1}^{2} r_k \sigma_k(J)(w_1^k - Jh) > \frac{\sum_{k=1}^{2} r_k \sum_{i=1}^{J} \sigma_k(i)(w_1^k - ih)}{J+1} > \sum_{k=1}^{2} r_k \sigma_k(J+1)(w_1^k - (J+1)h).$$

If, instead

$$c < \frac{1}{J^*+1} \sum_{k=1}^{2} r_k \sum_{i=1}^{J^*} \sigma_k(i)(w_1^k - ih). \tag{15}$$

then the solution will not be interior: an optimal depth-J knowledge hierarchy has $z_0 = 0$ and $\overline{z}_{J^*} = 1$.

The condition (15) is associated with a "complete" bureaucracy, in the sense of being capable of solving the entire range of problems. Completeness occurs when training and communication costs are low, partisanship is low, and Party 2's value of public goods is high.

5 State Capacity and Specialization

This section provides comparative statics results on state capacity and specialization as a function of the political environment. To begin, Figure 4 illustrates how state structures

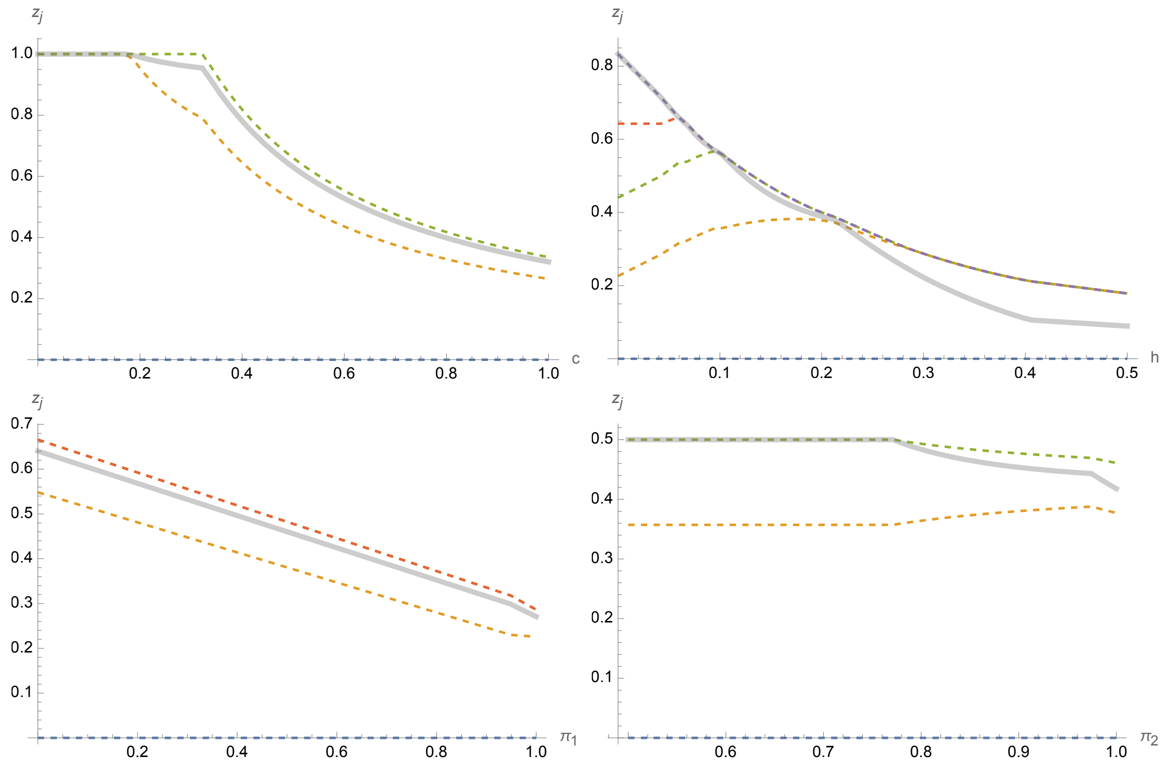


Figure 4: Optimal Knowledge Hierarchies and Output. Panels plot layers of optimal knowledge hierarchy established by Party 1, as a function of c, h, and π_k. Default parameters are $c = 0.7$, $h = 0.15$, $\pi_1 = 0.5$, $\pi_2 = 0.75$, $q = 0.5$, $\delta = 0.5$, $r_k = 0.5$, and $\epsilon = 0.25$. Dashed lines indicate layer boundaries, and gray solid lines indicate average output over time.

vary with some parameters of the model. It plots both Party 1's optimal knowledge sets, as well as average output after accounting for idling induced by realized election probabilities. An immediate observation in each example is that layers are occasionally idled, and thus output often falls short of capacity.

The figure shows that costs and partisanship have intuitive effects on the state's potential and realized problem-solving ability. In particular, higher costs and partisanship generally reduce capacity. The following comparative statics on $\overline{z}_{J^*}$ follow directly from expression (14), and so we state them without proof.

Remark 2 [Partisanship, Personnel Costs, and Capacity.] *At the optimal depth J^*, $\overline{z}_{J^*}$ is decreasing in π_k, c, and h.*

Figure 4 shows both the patterns in Remark 2, and that capacity and specialization are

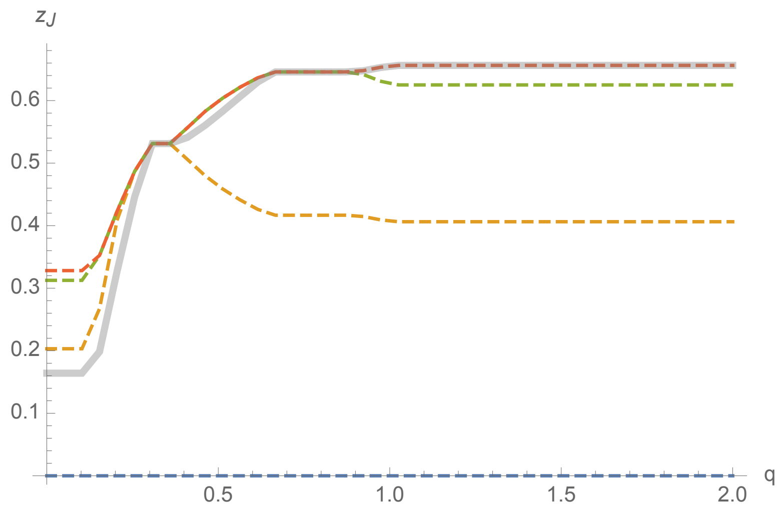


Figure 5: Insulation of state capacity from idling. Parameters are $\pi_1 = 0$, $c = 0.8$, $h = 0.15$, $\pi_2 = 0.75$, $r_k = 0.5$, $\delta = 0.5$, and $\epsilon = 0.25$.

not necessarily correlated. In particular, the cost parameters c and h have different implications for specialization. Along with reducing capacity, higher personnel costs generate greater depth. This contrasts with the effect of increasing operational/communication costs (h), which reduces both depth and capacity. High personnel costs increase the gains from specialization, while high operational costs reduce them.

Comparative statics with respect to q and r are less straight-forward, but show a number of interesting responses to partisanship. We illustrate this with a single example in Figure 5, before turning to a more complete exploration in the next two subsections. When the opponent's public good valuation q is very low, they receive little benefit from the state whether they are in or out of power—so they will always idle the entire state when they get the chance. As such, the state designer chooses an optimally specialized arrangement of layers for a low level of capacity, as shown in the left-hand side of the figure. As q increases the opponent is willing to keep a single layer open (sometimes), and the state designer capitalizes on this by expanding the state, while reducing the number of layers, so that more capacity will be "insulated" from politics. That is, by bundling together enough problems, the state designer ensures greater, although less efficient, state output. As q continues to increase, the opponent becomes willing to keep more layers open, and so the state designer increases both the size and complexity of the state. Finally, as q approaches one, both groups are willing to keep the same number of layers open, and so state capacity and output stabilize at a high level (indeed, higher than the opponent would like, as we show in Section 6).

5.1 Public Goods Valuation

Both capacity and specialization may be increasing or decreasing in q, as shown in Figure 6. They coincide when both parties share a consensus on the value of government production (q close to 1), but otherwise the prospect of idling creates a complex set of tradeoffs for the institution designer. Figure 6 strongly suggests that the effect of q on capacity depends on Party 1's partisanship (π_1). When π_1 is low, Party 1 loses little from being out of power. An opposition with a high public goods valuation will tend to activate layers and thereby induce the creation of a larger state. When π_1 is high, Party 1 does not benefit from the opposition's operation of the state, and compensates in part by reducing training costs. (Figure 6 also shows that this may entail reducing the state's depth.) Thus, higher values of q can reduce capacity.

Proposition 3 formalizes the relationship between q, π_1, and state capacity by showing that $\overline{z}_{J^*}$ (given by expression (14) decreases in q under high partisanship, and increases otherwise. The result makes use of the fact that J is integer-valued, and thus the optimal depth of the state is locally constant with respect to q.

Proposition 3 [Opposition Characteristics and Capacity.] *At the optimal depth J^*, there exists $\tilde{\pi}_1$ such that $\overline{z}_{J^*}$ is weakly decreasing (increasing) in q if $\pi_1 > (<)\ \tilde{\pi}_1$.*

The most interesting feature of Figure 6 is the non-monotonic relationship between capacity and the underlying number and arrangement of layers. Specialization can both increase and decrease with the state's potential problem-solving ability. These patterns are due to politics, as Party 1's design anticipates idling. As an example, consider the case where Party 1's partisanship is very low. At the lowest values of q, Party 2 idles all layers with certainty, and Party 1 designs a small state that is optimally specialized for a world in which only it participates in governance. As q increases and Party 2 becomes willing to

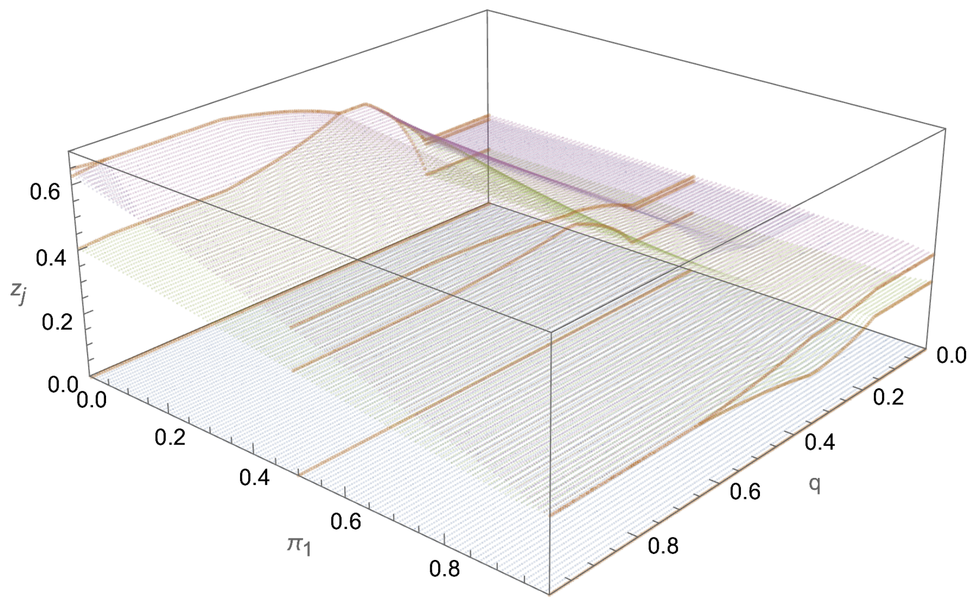


Figure 6: Optimal Knowledge Hierarchies as a Function of q and π_1. Figure plots layers of optimal knowledge hierarchy established by Party 1 at different values of π_1. Parameters are $c = 0.8$, $h = 0.15$, $\pi_2 = 0.75$, $\delta = 0.5$, $r_k = 0.5$, and $\epsilon = 0.25$.

activate the less expensive lower layers, Party 1 dispenses with upper layers that Party 2 is more likely to idle, thus reducing specialization in favor of greater capacity. Finally, when q is high enough so that Party 2 usually activates all layers, a specialized, high capacity state emerges.

Propositions 4 and 5 address Party 1's incentive to expand or contract layers as a function of its own partisanship and the opposition's public goods valuation. These parameters are central to the result because they determine the expected value of additional layers. Formally, the result examines the behavior of overall capacity $\overline{z}_{J^*}$, in relation to the minimum capacity levels that sustain J^* and J^*+1 layers (respectively $\underline{z}_{J^*}$ and $\underline{z}_{J^*+1}$, as given by expressions (12) and (13)). If $\overline{z}_{J^*} - \underline{z}_{J^*}$ is increasing and $\underline{z}_{J^*+1} - \overline{z}_{J^*}$ is decreasing in q, then the knowledge hierarchy moves closer to admitting a J^*+1th layer. In this case, we say that the knowledge hierarchy is *expanding*. Similarly, if $\overline{z}_{J^*} - \underline{z}_{J^*}$ is decreasing and $\underline{z}_{J^*+1} - \overline{z}_{J^*}$ is increasing in q, then the knowledge hierarchy moves closer to losing its J^*th layer. In this case, we say that the knowledge hierarchy is *contracting*.

Proposition 4 [Specialization and Capacity (low q).] *(i) If* $q < \frac{h}{1-\pi_2(r_1-\varepsilon)}$, *then* $\overline{z}_{J^*}$, $\underline{z}_{J^*}$, *and* $\overline{z}_{J^*+1}$ *are constant in* q.

(ii) If $q \in \left(\frac{h}{1-\pi_2(r_1-\varepsilon)}, \frac{Jh}{1-\pi_2(r_1-\varepsilon)}\right]$, *then there exists* $\underline{\pi}_1$ *such that the knowledge hierarchy is:*

$$\begin{cases} \text{contracting, with } \overline{z}_{J^*} \text{ increasing in } q & \text{if } \pi_1 < \underline{\pi}_1 \\ \text{expanding, with } \overline{z}_{J^*} \text{ decreasing in } q & \text{if } \pi_1 > \underline{\pi}_1. \end{cases}$$

Proposition 4 shows that when q is such that Party 2 never activates the top layer of the knowledge hierarchy, specialization and capacity are inversely related. When Party 1 is not very partisan it uses low specialization to take advantage of Party 2's willingness to activate the lowest layers. Thus, a large but "bottom heavy" state that insulates basic tasks from politics emerges. The inefficiently large lower layer(s) contract the upper layers and may eventually lead to their elimination, as Figure 5 showed.

Under high partisanship, the reverse occurs: Party 1 shifts tasks to upper layers that Party 2 is less likely to activate. It also shrinks capacity to offset the increased costs. Instead of insulating tasks from politics, the knowledge hierarchy makes tasks inaccessible to the opposition with a "top heavy" state with more capacity in the upper layers. Large and costly upper layers increase the need for specialization, and thus depth may increase even as overall capacity declines. Figure 7 depicts the transition from contraction to expansion as partisanship increases.

Proposition 5 next shows that the results for state are almost reversed when q is high enough to induce Party 2 to activate all layers with positive probability.

Proposition 5 [Specialization and Capacity (high q).] *(i) If* $q \in \left(\frac{Jh}{1-\pi_2(r_1-\varepsilon)}, \frac{Jh}{1-\pi_2(r_1+\varepsilon)}\right)$ *then there exist* $\hat{\pi}_1$, $\hat{\underline{\pi}}_1$ *where* $\hat{\pi}_1 > \hat{\underline{\pi}}_1$ *such that the knowledge hierarchy is:*

$$\begin{cases} \text{expanding, with } \overline{z}_{J^*} \text{ increasing in } q & \text{if } \pi_1 < \hat{\underline{\pi}}_1 \\ \text{contracting} & \text{if } \pi_1 > \hat{\pi}_1. \end{cases}$$

(ii) If $q \geq \frac{Jh}{1-\pi_2(r_1+\varepsilon)}$, *then* $\overline{z}_{J^*}$, $\underline{z}_{J^*}$, *and* $\overline{z}_{J^*+1}$ *are constant in* q.

Coupled with low partisanship by the state designer, higher values of q imply greater political consensus. This produces an environment where greater capacity complements

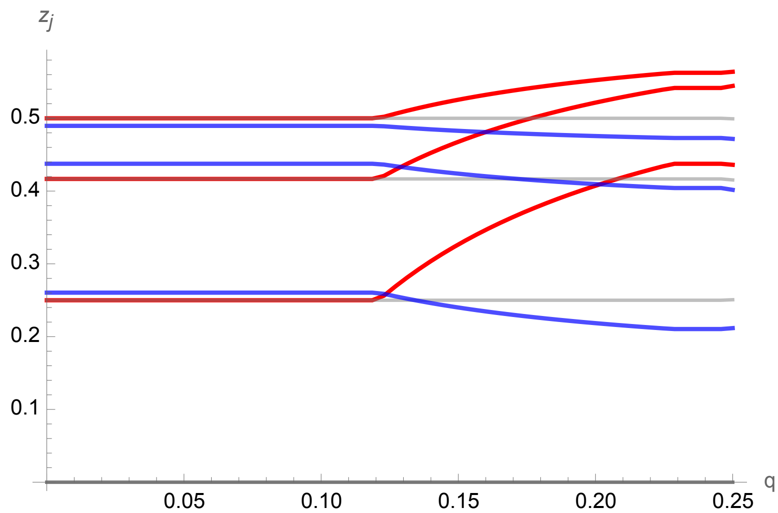


Figure 7: Expansion, Contraction, and Partisanship. Figure plots layers of optimal knowledge hierarchy established by Party 1, as a function of q. Parameters are $c = 0.6$, $h = 0.1$, $\pi_2 = 0.75$, $\delta = 0.5$, $r_k = 0.5$, and $\epsilon = 0.25$. Values of π_1 are 0.6 (in red), 0.9 (equal to $\tilde{\pi}_1$, in gray), and 0.98 (in blue).

greater specialization, and the state is an efficient problem-solver.[5] High partisanship now leads to contraction, though the effects on capacity are ambiguous. Finally, at the highest values of q, Party 2 activates layer J^* with certainty and all layers are constant in q until it becomes willing to activate a possible layer $J^* + 1$.

5.2 Transitions of Power

One sensible conjecture is that the ability to hold power will increase Party 1's return from state-building (e.g., Besley and Persson 2009). This logic implies that state capacity should increase in r_1. Our model produces this effect as well, but the possibility of idling sometimes changes this finding.

The prospects for holding power matter for the state designer when the stakes of losing power are high. Therefore, a highly partisan Party 1 invests in greater capacity as r_1 increases. Party 1 also creates greater capacity if Party 2 never activates any layers due to its own partisanship or low public goods valuation, or high operational costs. However, as partisanship declines and Party 2 becomes willing to activate layers, Party 1 benefits from

[5]High Party 2 partisanship (π_2) can undermine this, but not as easily as low public goods valuation.

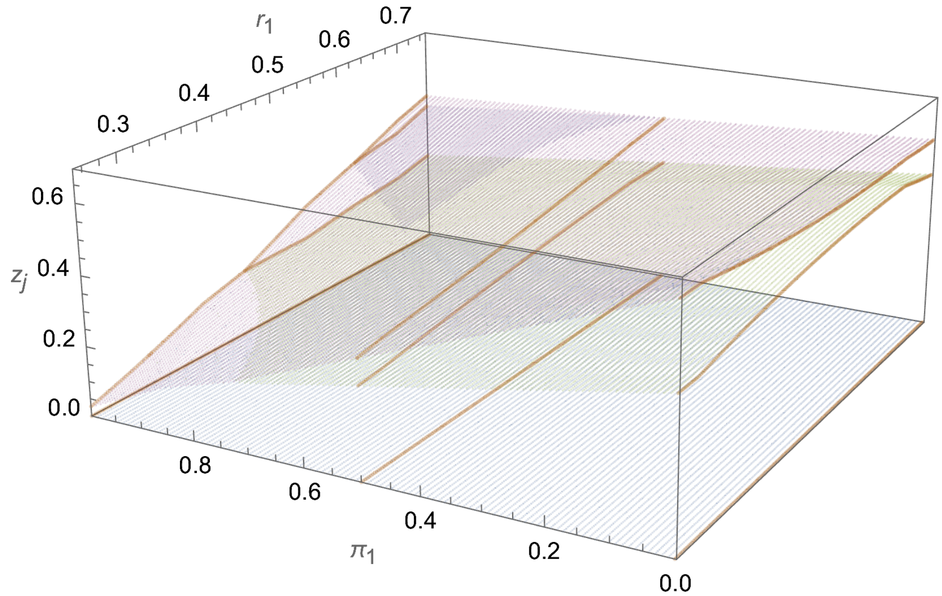


Figure 8: Optimal Knowledge Hierarchies as a Function of r_1 and π_1. Figure plots layers of optimal knowledge hierarchy established by party 1 at different values of π_1. Parameters are $c = 0.8$, $h = 0.15$, $\pi_2 = 0.75$, $q = 0.5$, $\delta = 0.5$, and $\epsilon = 0.25$.

the state even when out of power. This reduces the importance of holding power and and thus also its effect on capacity-building.

The next proposition establishes this relationship. A subtlety of the derivation is the fact that r_1 enters Party 1's objective through both directly through its probability of holding power and indirectly through each layer's long run probability of activation. For reduced partisanship to attenuate the relationship between r_1 and capacity, decreasing partisanship must also not encourage Party 1 to activate layers more often, which would amplify the effect of r_1. This condition holds when conditions permit Party 1 to activate the top layer with certainty.

Proposition 6 [Elections and Capacity.] *If* $\pi_1 > 1 - h$ *or* $h > q(1 - \pi_2(r_1 - \varepsilon))$, *then at the optimal depth* $\overline{z}_{J^*}$ *is weakly increasing in* r_1. *If* $\pi_1 < (1 - hJ^*)/(1 - (r_1 - \varepsilon))$, *then* $\frac{\partial \overline{z}_{J^*}}{\partial r_1}$ *is weakly increasing in* π_1.

Figure 8 shows how reducing Party 1's partisanship generally increases state capacity and flattens the effect of r_1. Interestingly, it shows that the relationship between capacity and the probability of holding power not only weakens but reverses as partisanship decreases. To understand why this reversal can occur, it is again important to take Party 2's idling incentives into account. Under the parameters assumed in the figure, Party 2 activates layer 1 with probability $2.37 - 2r_1$ and layer 2 with probability $1.57 - 2r_1$. Thus as Party 1 increasingly values public goods when Party 2 is in power, it shifts more problem-solving capacity to the first layer, which is less likely to be idled. A low-capacity state — driven by the a shrinkage of the second layer—compensates for the high personnel costs incurred by a very large first layer.

6 Counterfactuals

The political process often produces a state that looks substantially different from one that is designed to maximize some notion of welfare. Imperfect state maintenance reduces welfare in and of itself, but also can produce incentives to create states that are of sub-optimal size—either too large or too small, and complexity—too complex, or too simple. In this section, we characterize some of these distortions and show that whether a state is sub-optimally large or small depends crucially on q, the public goods valuation of group 2.

In order to focus on the distortions due to sub-optimal state capacity and imperfect state maintenance, we specify welfare maximization in a particular way that ignores the effects of sub-optimal spending decisions. There are effectively three ways in which the political process distorts outcomes away from the welfare maximizing outcome: imperfect state maintenance, sub-optimal state design, and sub-optimal spending decisions. The last option, in particular has two facets. First, welfare might be maximized by choosing to spend state revenues on things other what either party might spend it on. We have ruled out this possibility in the model up until this point, and we continue to rule it out now. Even so, one of those spending choices will create higher welfare than the other. For example, if r_1—the size of group 1, and its probability of being in power—is close to 1, then it will maximize welfare to always spend money on the same thing group 1 would, if it were in power.

We ignore the political distortions to welfare due to sub-optimal policy choice by considering a social planner (SP) that will spend money on the priorities of each group with the same probability as the political process would generate. In particular, the SP will spend state revenues in the same way as as party k with probability r_k. Thus, the expected welfare from state spending in a given period is

$$w_{SP} = \sum_{k=1}^{2} r_k \sum_{i=1}^{2} r_i w_i^k$$

In this formulation, r_i reflects the proportion of the population in each group. The SP will also take into account the costs of generating state revenue. Finally, we assume the SP must always run the state at full capacity so that they do not, for example, idle the entire state when r_1 is close to 1, and it looks likely that in the next period the SP will have to spend money in the same way as group 2.

The social planner will usually produce different state structures—in particular capacity and structure—than Party 1 does in the political process. We illustrate this by varying q in Figure 9 using many of the same parameters as in Figure 6, focusing on the cross-sections when $\pi_1 = 0$ (on the left-hand side) and $\pi_1 = 1$ (on the right-hand side). In the first of these cross sections, under the political process state capacity and output are increasing in q, although the complexity of the state is non-monotonic, as predicted by Propositions 4 and 5. In the latter cross-section, capacity and complexity are decreasing, although output is non-monotonic. The optimal structure under the social planner is much more straight-forward: in both cross-sections capacity, complexity, and output are monotonically increasing. These latter relationships make sense: in both panels, w_{SP} is strictly increasing in q, which increases the value of state capacity. Without distortions introduced by imperfect state maintenance, this increasing value translates into increasing capacity, with increasing complexity reducing the aggregate cost of output.

An obvious pattern emerges by comparing the patterns of state capacity and output in Panel C of Figure 9: when the party that values public goods more designs the state ($q < 1$) it tends to be sub-optimally large, whereas when the party that values public goods less ($q > 1$) designs the state, it will tend to be sub-optimally small. However, the distortions in state maintenance, and therefore complexity, will tend to only occur when the party designing the state values public goods more, as can be seen from comparing the red to

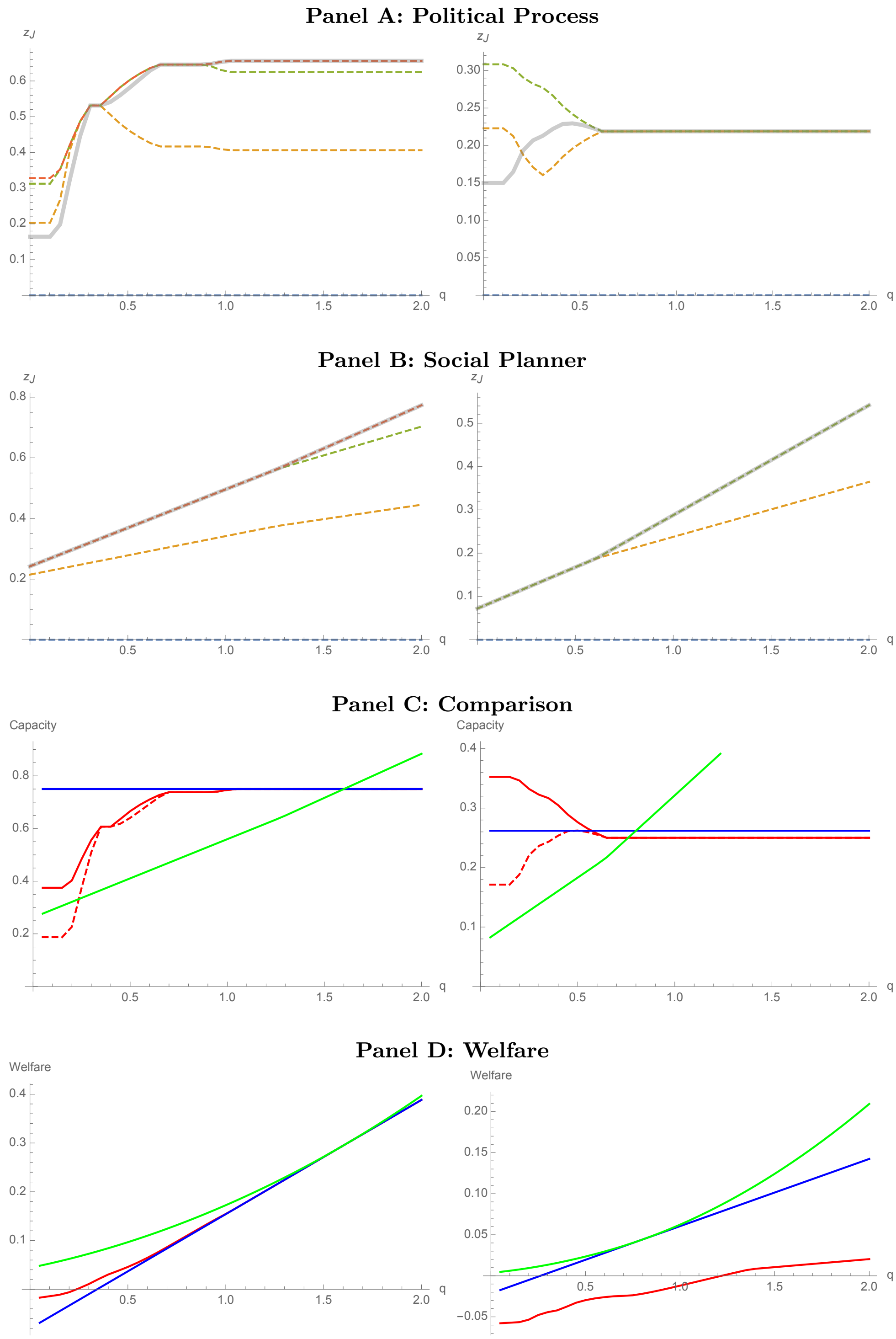


Figure 9: Optimal Knowledge Hierarchies for Political Process and Social Planner. In bottom panels, red lines are state capacity and output under political processes, blue is without the possibility of idling layers, and green is the social planner. Default parameters are $c = 0.8$, $h = 0.15$, $\pi_2 = 0.75$, $r_k = 0.5$, $\delta = 0.5$, and $\epsilon = 0.25$. Left-hand panels have $\pi_1 = 0$, right-hand panels $\pi_1 = 1$.

green lines in Panel C.[6] This makes intuitive sense: when the party designing the state values public goods less, the other party will tend to be more interested in keeping layers active. With no distortions in state maintenance, political distortions to complexity are unnecessary. We can formalize this pattern and intuition.

Proposition 7 *If, in the neighborhood of* $q = \frac{1-r_2\pi_1}{1-r_1\pi_2}$, J^* *is the same under both the social planner's and political problem, and* $\sigma_k(j) = 1$ *for* $j \leq J$, *then the state will be sub-optimally large under the political problem if and only if* $q > \frac{1-r_2\pi_1}{1-r_1\pi_2}$.

An obvious case to examine is when $r_k = 1/2$ and partisanship is symmetric ($\pi_1 = \pi_2$). In this case, state capacity will be sub-optimally large if and only if $q < 1$, and sub-optimally small otherwise.

7 Discussion

Despite a recent surge in scholarly interest in the topic, theoretical accounts of state capacity have taken little notice of several key aspects of governmental functions and organizations. In addition to revenue collection, modern governance requires the skilled performance of a wide range of activities, each of which presents distributions of sub-problems of varying levels of difficulty. Modern states are also typically subject to regular political processes. Politicians with heterogeneous preferences may interfere with its functioning, and recent decades have seen the emergence of numerous political movements that are expressly interested in reducing the power of the state. Everyday state maintenance is therefore important even after the critical moment of state foundation.

Our theory of state capacity posits the bureaucracy as the primary link between politicians and the solution of social problems. Knowledge hierarchies provide a flexible and tractable way to model the throughput of such complex organizations. Existing applications focus on the structure of firms, but its natural trade-offs between specialization and coordination pervades public sector organizations as well. By placing knowledge hierarchies in a simple dynamics political economy setting, our model provides a novel application for the framework.

[6] Which is why we have focused on $q < 1$ up until this section.

The ability of politicians to idle parts of the state allows the model to distinguish meaningfully between output and capacity. Incumbent politicians idle layers of the state according to projected election outcomes and policy preferences, and political competitiveness does not minimize the risk of idling. These idling strategies in turn produce distortions in the capacity and specialization dimensions of state design. Most notably, high capacity and specialization coincide only when social groups are not polarized and agree on the value of policy. When the founding party or group faces an opposition that cares less about the policy in question, this relationship is broken: higher capacity and less specialization, or lower capacity and greater specialization can result, depending on the founder's partisanship. Partisanship also affects the relationship between capacity and electoral prospects. An electorally advantaged founding group maximizes capacity under high partisanship, but low partisanship it is an electorally disadvantaged group that may do so.

Our framework presents numerous directions for additional work. In addition to providing results on state structure, the model should provide observable implications on realized output. There is also room for exploring robustness with respect to policy or organizational technologies. For example, groups may have different valuations over specific problems, as opposed to over policy production in general. As the IRS example in the introduction illustrates, liberal and conservative groups might prefer to idle different parts of the state because of their activities affect core constituents. Alternatively, some organizations may feature top-to-bottom propagation of problems, rather than bottom-to-top. It is finally worth considering in general how the technology of knowledge hierarchies can fit into other political or institutional contexts.

8 Appendix

Proof of Proposition 1. The ex ante probability that layer j is active in any given period is $\sum_k r_k \sigma_k(j)$. Ignoring corner solutions, for parties 1 and 2, the conditional probability $\sigma_k(j) \equiv 1 - F_k(p_k^j)$ of keeping a layer active is:

$$\begin{aligned}
\hat{\sigma}_1(j) &= \frac{1}{2\varepsilon}\left[r_1 + \varepsilon - \frac{jh - w_1^2}{2\varepsilon(w_1^1 - w_1^2)}\right] = \frac{(r_1+\varepsilon)\pi_1 - (jh - (1-\pi_1))}{2\varepsilon\pi_1} \\
\hat{\sigma}_2(j) &= \frac{1}{2\varepsilon}\left[r_2 + \varepsilon - \frac{jh - w_2^1}{2\varepsilon(w_2^2 - w_2^1)}\right] = \frac{(1-r_1+\varepsilon)q\pi_2 - (jh - q(1-\pi_2))}{2\varepsilon q\pi_2}.
\end{aligned}$$

Noting that $r_1 = 1 - r_2$, each $r_k\hat{\sigma}_k(j)$ is convex in r_1. Moreover, it is apparent that for $r_1 \in [0,1]$, $\hat{\sigma}_1(j) > 0$ only if r_1 is sufficiently high, and $\hat{\sigma}_2(j) > 0$ only if r_1 is sufficiently low. Let $\underline{r}_k^j$ denote the value of r_1 such that $\sigma_k(j) = 0$, and $\overline{r}_k^j$ denote the value of r_1 such that $\sigma_k(j) = 1$, if such values exist. Then we have the following:

$$\sigma_1(j) = \begin{cases} 0 & \text{if } r_1 \le \underline{r}_1^j \\ \hat{\sigma}_1(j) & \text{if } r_1 \in (\underline{r}_1^j, \overline{r}_1^j) \\ 1 & \text{if } r_1 \ge \overline{r}_1^j \end{cases} \qquad \sigma_2(j) = \begin{cases} 1 & \text{if } r_1 \le \overline{r}_2^j \\ \hat{\sigma}_2(j) & \text{if } r_1 \in (\underline{r}_2^j, \overline{r}_2^j) \\ 0 & \text{if } r_1 \ge \underline{r}_2^j. \end{cases} \tag{16}$$

We now establish (i). First, $\hat{\sigma}_1(j)$ and $\hat{\sigma}_2(j)$ are both weakly decreasing in h, thus, $\sum_k r_k\sigma_k(j)$ is also weakly decreasing in h. Second, $\hat{\sigma}_2(j)$ is weakly increasing in q, and $\hat{\sigma}_1(j)$ is constant in q, thus, $\sum_k r_k\sigma_k(j)$ is weakly increasing in q. Third, taking derivatives with respect to π_k produces:

$$\frac{d\hat{\sigma}_1(j)}{d\pi_1} = -\frac{1-jh}{2\varepsilon\pi_1^2} \qquad \frac{d\hat{\sigma}_2(j)}{dw_2} = -\frac{q-jh}{2\varepsilon q\pi_2^2}.$$

This implies that the $\hat{\sigma}_k(j)$ are monotonic in π_k on $[0,1]$. Further, if $\frac{d\hat{\sigma}_1(j)}{d\pi_1} > 0$, then $jh > 1$. As $\hat{\sigma}_1(j)$ is maximized when $r_1 + \varepsilon = 1$, $\frac{d\hat{\sigma}_1(j)}{d\pi_1} > 0 \Rightarrow \hat{\sigma}_1(j) < 0$, and thus $\sigma_1(j) = 0$ (a similar argument works for $\sigma_2(j)$). Thus, $\hat{\sigma}_k(j)$ must be weakly decreasing in π_k, which implies that $\sum_k r_k\sigma_k(j)$ is also weakly decreasing in π_k.

(ii) $\sum_k r_k\sigma_k(j) = 0$ iff the interval $[\underline{r}_2^j, \underline{r}_1^j]$ is non-empty. Solving for the value of r_1 such that $\hat{\sigma}_1(j) = 0$ produces $\underline{r}_1^j = 1 - \frac{1-jh}{\pi_1} - \varepsilon$. Similarly, solving for the value of r_1 such that

$\hat{\sigma}_2(j) = 0$ produces $\underline{r}_2^j = \frac{q-jh}{q\pi_2} + \varepsilon$. As $\underline{r}_2^j$ is increasing in q, the interval is nonempty if $q < \frac{jh\pi_1}{\pi_1+\pi_2((1-jh)-\pi_1(1-2\varepsilon))}$. For this interval to be internal both $\underline{r}_1^j < 1$ and $\underline{r}_2^j > 0$. The fact that the first holds is implied by the layer's existence. The second holds if $jh < q(1+\varepsilon\pi_2)$.

(iii) $\sum_k r_k\sigma_k(j) = 1$ iff the interval $[\overline{r}_1^j, \overline{r}_2^j]$ is non-empty. Solving for the value of r_1 such that $\hat{\sigma}_1(j) = 1$ produces $\overline{r}_1^j = 1 - \frac{1-jh}{\pi_1} + \varepsilon$. Similarly, solving for the value of r_1 such that $\hat{\sigma}_2(j) = 1$ produces $\overline{r}_2^j = \frac{q-jh}{q\pi_2} - \varepsilon$. Since $\overline{r}_2^j$ is increasing in q, the interval is nonempty if $q > \frac{jh\pi_1}{\pi_1+\pi_2((1-jh)-\pi_1(1+2\varepsilon))}$. For this interval to be internal both $\overline{r}_1 > 0$ and $\overline{r}_2 < 1$. The first will hold if $jh > 1 - \pi_1(1+\varepsilon)$, and the second if $jh > q(1-\pi_2(1+\varepsilon))$, so both will hold when $jh > \max[1-\pi_1(1+\varepsilon), q(1-\pi_2(1+\varepsilon))]$. ∎

Proof of Lemma 1. We prove the result by evaluating the ex-ante expected payoff to a member of Party 1 from the knowledge hierarchy.

Note that if $[z_j, \overline{z}_j] \cap [z_{j'}, \overline{z}_{j'}] \neq \emptyset$, for some layer j and j'—that is, two layers overlap—then the party could do strictly better by eliminating all workers in the intersection in either layer j or j'. Thus an optimal knowledge hierarchy must have non-overlapping layers.

Recall that Party 1 designs the knowledge hierarchy in period 0. Consider two successive layers with a gap between them, that is, $Z_j = [\overline{z}_{j-1} + \eta, \overline{z}_j]$, for some $\eta \in (0, \overline{z}_j - \overline{z}_{j-1})$. The ex-ante expected value of a given layer j is:

$$\left[\sum_{t=1}^{\infty} \delta^t \sum_{k=1}^{2} r_k\sigma_k(j)(w_1^k - jh) - c\overline{z}_j\right](G(\overline{z}_j) - G(\overline{z}_{j-1} + \eta)) > 0$$

which is positive for any layer that is created in equilibrium. Denoting the term in square braces by ϕ_j, we have $\phi_j > 0$ in equilibrium as $\overline{z}_j > \overline{z}_{j-1} + \eta$. The derivative of the above expression with respect to η is then $-g(\overline{z}_{j-1} + \eta)\phi_j$, which is negative as $g(\cdot)$ has full support. Thus, Party 1 will want to set $\eta = 0$, leading to no gaps. The fact that $z_1 = 0$ follows from the same argument applied to $z_1 = \eta$. ∎

Proof of Lemma 2. There are two cases. In the first, $z_0 = 0$ and $\overline{z}_J < 1$. The Hessian of (9) with respect to $\overline{z}_1, \ldots, \overline{z}_J$ is $\frac{\delta}{1-\delta}$ times the matrix

$$\overline{z}_1 \quad \overline{z}_2 \quad \overline{z}_3 \quad \ldots \quad \overline{z}_{J-1} \quad \overline{z}_J$$

$\overline{z}_1$	$-2c$	c	0	0	0	0
$\overline{z}_2$	c	$-2c$	c	0	0	0
$\overline{z}_3$	0	c	$-2c$	$\ldots$	0	0
$\vdots$	0	0	$\ldots$	$\ldots$	$\ldots$	0
$\overline{z}_{J-1}$	0	0	0	$\ldots$	$-2c$	c
$\overline{z}_J$	0	0	0	0	c	$-2c$

Thus, the Hessian is negative definite for any $c > 0$. ■

Proof of Proposition 2. Let $U_1^*(z^s)$ be the Party 1 politician's utility over such knowledge hierarchies at the optimal depth for a given capacity z^s. Our first objective is to show that $U_1^*(z^s)$ is concave and differentiable. We construct $U_1^*(z^s)$ piecewise by determining the expected value of an optimal knowledge hierarchy at the optimal depth J^* for each possible capacity level z^s. As argued in the text, the optimal depth is the maximum number of layers such that each has non-negative length.

Let $U_1^*(\overline{z}_J, J)$ denote the politician's objective for given values of $\overline{z}_J$ and J, where all layer boundaries below J are arranged optimally. By Lemma 1, $\overline{z}_J$ is also the capacity of the knowledge hierarchy. We can then rewrite the objective (9) in terms of $\overline{z}_j$ as follows:

$$U_1^*(\overline{z}_J, J) = \frac{\delta}{1-\delta} \sum_{j=1}^{J} \left[\sum_{k=1}^{2} r_k \sigma_k(j)(w_1^k - jh) - c\overline{z}_j \right] (\overline{z}_j - \overline{z}_{j-1}) \, .$$

$U_1^*(\overline{z}_J, J)$ is weakly increasing in J, since a politician can do as well as a depth-J knowledge hierarchy with a depth-$J+1$ knowledge hierarchy with a degenerate $J+1$th knowledge set of length 0.

To complete the expression of $U_1^*(\overline{z}_J, J)$ we write each $\overline{z}_j$ in terms of z_0 and $\overline{z}_J$. To reiterate from (11), each $\overline{z}_j$ can be written as follows:

$$\overline{z}_j \quad = \quad \frac{j}{J}\overline{z}_J + \frac{1}{cJ} \sum_{k=1}^{2} r_k \left(J \sum_{i=1}^{j} \sigma_k(i)(w_1^k - ih) - j \sum_{i=1}^{J} \sigma_k(i)(w_1^k - ih) \right) .$$

It will be convenient to denote the last part of the preceding expression as follows:

$$\kappa_{j,J} = \frac{1}{cJ}\sum_{k=1}^{2} r_k \left(J\sum_{i=1}^{j}\sigma_k(i)(w_1^k - ih) - j\sum_{i=1}^{J}\sigma_k(i)(w_1^k - ih)\right)..$$

Observe that $\kappa_{j,J}$ is independent of all $\overline{z}_j$, and $\kappa_{0,J} = \kappa_{J,J} = 0$.

As noted in the text, expression (13) gives $\underline{z}_J$, which is the minimum capacity necessary to sustain J layers of non-negative length. As defined, $U_1^*(\underline{z}_J, J) = U_1^*(\underline{z}_J, J-1)$. We can therefore construct $U_1^*(z^s)$ as a continuous function as follows:

$$U_1^*(z^s) = U_1^*(\overline{z}_J, J) \quad \text{for } z^s \in \left[\underline{z}_J, \underline{z}_{J+1}\right).$$

We now show that $U_1^*(z^s)$ is differentiable and concave using first- and second-order conditions. Omitting terms that are independent of $\overline{z}_j$ and using the fact that $z_0 = 0$, the politician's objective for a given J simplifies to:

$$U_1^*(\overline{z}_J, J) = \frac{\delta}{1-\delta}\sum_{j=1}^{J}\left[\sum_{k=1}^{2}\left(r_k\sigma_k(j)(w_1^k - jh) - c\left(\frac{j\overline{z}_J}{J} + \kappa_{j,J}\right)\right)\left(\frac{\overline{z}_J}{J} + \kappa_{j,J} - \kappa_{j-1,J}\right)\right].$$

Differentiating with respect to $\overline{z}_J$ produces the following expression, which we designate by parts:

$$\frac{dU_1^*}{d\overline{z}_J} = \sum_{j=1}^{J}\left[\underbrace{\left(\sum_{k=1}^{2}\frac{r_k\sigma_k(j)(w_1^k - jh)}{J}\right)}_{A} - \underbrace{\frac{2cj\overline{z}_J}{J^2}}_{B} - \underbrace{\frac{c}{J}\left((j+1)\kappa_{j,J} - j\kappa_{j-1,J}\right)}_{C}\right], \tag{17}$$

in which part 'C' equals zero:

$$\sum_{j=1}^{J}\frac{c}{J}\left((j+1)\kappa_{j,J} - j\kappa_{j-1,J}\right) = \frac{c}{J}\left((J+1)\kappa_{J,J} + \kappa_{0,J}\right) = 0. \tag{18}$$

We now show that:

$$\frac{dU_1^*(\underline{z}_J, J-1)}{d\overline{z}_J} = \frac{dU_1^*(\underline{z}_J, J)}{d\overline{z}_J}.$$

To show this, we compare differences between the values of each part at J and $J-1$ layers.

For part 'A' of (17) we have for $J-1$ and J, respectively:

$$\begin{aligned}
\sum_{j=1}^{J-1}\sum_{k=1}^{2}\frac{r_k\sigma_k(j)(w_1^k-jh)}{(J-1)} &= \sum_{j=1}^{J-1}\sum_{k=1}^{2}\frac{r_k\sigma_k(j)(w_1^k-jh)J}{J(J-1)} \\
\sum_{j=1}^{J}\sum_{k=1}^{2}\frac{r_k\sigma_k(j)(w_1^k-jh)}{J} &= \sum_{j=1}^{J}\sum_{k=1}^{2}\frac{r_k\sigma_k(j)(w_1^k-jh)J}{J(J-1)} - \sum_{j=1}^{J}\sum_{k=1}^{2}\frac{r_k\sigma_k(j)(w_1^k-jh)}{J(J-1)}.
\end{aligned}$$

Taking the difference between the right-hand side terms produces:

$$\sum_{k=1}^{2}\left(\sum_{j=1}^{J-1}\frac{r_k\sigma_k(j)(w_1^k-jh)}{J(J-1)} - \frac{r_k\sigma_k(J)(w_1^k-Jh)(J-1)}{J}\right).$$

Next, for part 'B' of (17), we substitute in values of $\underline{z}_J$ (13) for $J-1$ and J, respectively:

$$\begin{aligned}
&\sum_{j=1}^{J-1}\frac{2cj}{(J-1)^2}\sum_{k=1}^{2}\frac{r_k}{c}\left(\sum_{i=1}^{J-1}\sigma_k(i)(w_1^k-ih)-(J-1)\sigma_k(J)(w_1^k-Jh)\right) = \\
&\qquad \frac{J}{J-1}\sum_{k=1}^{2}r_k\left(\sum_{i=1}^{J-1}\sigma_k(i)(w_1^k-ih)-(J-1)\sigma_k(J)(w_1^k-Jh)\right) \\
&\sum_{j=1}^{J}\frac{2cj}{J^2}\sum_{k=1}^{2}\frac{r_k}{c}\left(\sum_{i=1}^{J-1}\sigma_k(i)(w_1^k-ih)-(J-1)\sigma_k(J)(w_1^k-Jh)\right) = \\
&\qquad \frac{(J+1)}{J}\sum_{k=1}^{2}r_k\left(\sum_{i=1}^{J-1}\sigma_k(i)(w_1^k-ih)-(J-1)\sigma_k(J)(w_1^k-Jh)\right).
\end{aligned}$$

Taking the difference between these terms produces:

$$\begin{aligned}
&\frac{1}{J(J-1)}\sum_{i=1}^{J-1}\sum_{k=1}^{2}r_k\sigma_k(i)(w_1^k-ih)-(J-1)r_k\sigma_k(J)(w_1^k-Jh) = \\
&\qquad \sum_{k=1}^{2}\left(\sum_{i=1}^{J-1}\frac{r_k\sigma_k(i)(w_1^k-ih)}{J(J-1)} - \frac{(J-1)r_k\sigma_k(J)(w_1^k-Jh)}{J}\right).
\end{aligned}$$

Thus, the costs and benefits in 'A' and 'B' cancel. Thus $U_1^*(\overline{z}_J, J)$ and $U_1^*(\overline{z}_J, J-1)$ are tangent at $\underline{z}_J$, and so $U_1^*(z^s)$ is differentiable at $z^s = \underline{z}_J$.

We next show that each $U_1^*(\overline{z}_J, J)$ is concave in $\overline{z}_J$. The second derivative of $U_1^*(\overline{z}_J, J)$ with respect to $\overline{z}_J$ is $-2(J+1)c/J$, which is clearly negative. Thus $U_1^*(z^s)$ inherits concavity from the constituent $U_1^*(\overline{z}_J, J)$ functions.

What remains is characterizing the optimal capacity and depth. For any given J, the unique interior solution is derived by straightforward differentiation to produce expressions (11) and (12). Since $U_1^*(z^s)$ is concave, at an interior solution the maximizer $\overline{z}_{J^*}$ must coincide with the unique value of J such that $\overline{z}_{J^*} \in [\underline{z}_J, \underline{z}_{J+1})$. Otherwise, the optimal state is at a corner. If $\frac{dU_1^*(0)}{dz^s} < 0$ then $\overline{z}_{J^*} = 0$ and no state is optimal. And if $\frac{dU_1^*(1)}{dz^s} > 0$ then $\overline{z}_J^* = 1$ and the state solves all problems. Condition (15) is obtained by solving (12) for c at depth J^*.

To characterize J^* and obtain the expression in the statement of the result, observe that the optimal position of the last layer boundary, $\overline{z}_J$ (as given by (12)) must satisfy (i) $\overline{z}_J - \underline{z}_J \geq 0$; and (ii) $\underline{z}_{J+1} - \overline{z}_J \geq 0$. Substituting, conditions (i) and (ii) can be rewritten respectively as:

$$\sum_{k=1}^{2} \frac{r_k J \left[(J+1)(w_1^k - Jh)\sigma_k(J) - \sum_{i=1}^{J}(w_1^k - ih)\sigma_k(i)\right]}{(J+1)c} \geq 0 \quad (19)$$

$$\sum_{k=1}^{2} \frac{r_k J \left[\sum_{i=1}^{J}(w_1^k - ih)\sigma_k(i) - (J+1)(w_1^k - (J+1)h)\sigma_k(J+1)\right]}{(J+1)c} \geq 0. \quad (20)$$

Each of these expressions respectively hold iff:

$$\sum_{k=1}^{2} r_k(J+1)\sigma_k(J)(w_1^k - Jh) \geq \sum_{k=1}^{2}\sum_{i=1}^{J} r_k\sigma_k(i)(w_1^k - ih) \quad (21)$$

$$\sum_{k=1}^{2} r_k(J+1)\sigma_k(J+1)(w_1^k - (J+1)h) \leq \sum_{k=1}^{2}\sum_{i=1}^{J} r_k\sigma_k(i)(w_1^k - ih) \quad (22)$$

Since both $\sigma_k(i)$ and $w_1^k - ih$ are non-increasing in i, it is clear that each increment in J adds at most $\sum_k r_k\sigma_k(J)(w_1^k - Jh)$ and $\sum_k r_k\sigma_k(J+1)(w_1^k - (J+1)dh)$ to the left-hand sides of (21) and (22), respectively. It adds exactly $\sum_k r_k\sigma_k(J)(w_1^k - Jh)$ to their right-hand sides. Thus (19) holds only for J sufficiently small, and (20) holds only for J sufficiently large. J^* is the unique value of J satisfying both. ■

Proof of Proposition 3. Differentiating the expression for $\overline{z}_{J^*}$ (14) with respect to q produces:

$$\frac{\partial \overline{z}_{J^*}}{\partial q} = \frac{r_2}{c(J^*+1)} \sum_{i=1}^{J^*} (1 - \pi_1 - ih)\frac{\partial \sigma_2(i)}{\partial q}. \quad (23)$$

For a fixed J^*, this derivative will be positive iff $\sum_{i=1}^{J^*}(1-\pi_1-ih)\frac{\partial\sigma_2(i)}{\partial q} > 0$.

Observe that $\frac{\partial\sigma_2(i)}{\partial q} > 0$ at an interior solution, $\frac{\partial\sigma_2(i)}{\partial q} = 0$ at a corner solution, and $\frac{\partial\sigma_2(i)}{\partial q}$ is independent of π_1. Thus for any J^*, $\frac{\partial \overline{z}_{J^*}}{\partial q} > (=)(<)\ 0$ if:

$$\pi_1 < (=)(>)\ \tilde{\pi}_1 \equiv 1 - \frac{h\sum_{i=1}^{J^*} i\frac{\partial\sigma_2(i)}{\partial q}}{\sum_{i=1}^{J^*}\frac{\partial\sigma_2(i)}{\partial q}}. \tag{24}$$

The value of $\tilde{\pi}_1$ is clearly contained within $(1-Jh, 1)$. Note finally that in the special case where $\sigma_2(1), \ldots, \sigma_2(j') \in (0,1)$, this threshold simplifies to $1 - h(2j'+1)/3$. ■

Proof of Proposition 4. We prove the result by examining the behavior of $\overline{z}_J$ in relation to the minimum thresholds for J and $J+1$ layers. Define $z_- \equiv \overline{z}_J - \underline{z}_J$ for a given J. By the argument in the proof of Proposition 2, an optimal non-empty J-layer knowledge hierarchy is feasible if $z_- > 0$. A necessary condition for depth to decrease is $\frac{\partial z_-}{\partial q} < 0$.

Using expressions (12) and (13), z_- evaluates to:

$$z_- \;=\; \frac{J}{c(J+1)}\left(J\sum_{k=1}^{2} r_k\sigma_k(J)(w_1^k - Jh) - \sum_{k=1}^{2} r_k \sum_{i=1}^{J-1}\sigma_k(i)(w_1^k - ih)\right). \tag{25}$$

Differentiating z_- with respect to q produces:

$$\frac{\partial z_-}{\partial q} = \frac{Jr_2}{c(J+1)}\left(J(1-\pi_1-Jh)\frac{\partial\sigma_2(J)}{\partial q} - \sum_{i=1}^{J-1}(1-\pi_1-ih)\frac{\partial\sigma_2(i)}{\partial q}\right). \tag{26}$$

Now define $z_+ \equiv \underline{z}_{J+1} - \overline{z}_J$, again using (12) and (13). A necessary condition for depth to increase is $\frac{\partial z_+}{\partial q} < 0$.

$$z_+ = \frac{J}{c(J+1)}\left(\sum_{k=1}^{2} r_k \sum_{i=1}^{J}(w_1^k - ih)\sigma_k(i) - (J+1)\sum_{k=1}^{2} r_k(w_1^k - (J+1)h)\sigma_k(J+1)\right). \tag{27}$$

Differentiating z_+ with respect to q produces:

$$\frac{\partial z_+}{\partial q} = \frac{Jr_2}{c(J+1)}\left(\sum_{i=1}^{J}(1-\pi_1-ih)\frac{\partial\sigma_2(i)}{\partial q} - (J+1)(1-\pi_1-(J+1)h)\frac{\partial\sigma_2(J)}{\partial q}\right). \tag{28}$$

Next, note that under the assumption of uniformly distribution re-election probabilities we have at an interior solution:

$$\sigma_2(i) = \frac{1}{2} + \frac{r_2 q \pi_2 - jh + q(1-\pi_2)}{2\varepsilon q \pi_2}.$$

Observe that $\frac{\partial \sigma_2(i)}{\partial q} > 0$ at an interior solution, $\frac{\partial \sigma_2(i)}{\partial q} = 0$ at a corner solution, and $\sigma_2(i)$ is decreasing in i.

There are two cases, depending on the values of $\sigma_2(i)$.

(i) When $\sigma_2(i) = 0$ for all i (for which $\sigma_2(1) = 0$, or equivalently $q \leq \frac{h}{1-\pi_2(r_1-\varepsilon)}$, is sufficient), $\frac{\partial z_-}{\partial q}$ and $\frac{\partial z_+}{\partial q}$ are obviously zero. Thus there will be no changes in J^* until q increases enough such that $\frac{\partial \sigma_2(1)}{\partial q} > 0$.

(ii) Suppose $\sigma_2(J) = 0$ but $\sigma_2(i) > 0$ for some i, or equivalently $q \in \left(\frac{h}{1-\pi_2(r_1-\varepsilon)}, \frac{Jh}{1-\pi_2(r_1-\varepsilon)}\right]$. Then $\frac{\partial z_-}{\partial q} > 0$ iff:

$$\pi_1 > \underline{\pi}_1^- \equiv 1 - \frac{h \sum_{i=1}^{J-1} i \frac{\partial \sigma_2(i)}{\partial q}}{\sum_{i=1}^{J-1} \frac{\partial \sigma_2(i)}{\partial q}}.$$

Likewise, $\frac{\partial z_+}{\partial q} > 0$ iff:

$$\pi_1 < \underline{\pi}_1^+ \equiv 1 - \frac{h \sum_{i=1}^{J} i \frac{\partial \sigma_2(i)}{\partial q}}{\sum_{i=1}^{J} \frac{\partial \sigma_2(i)}{\partial q}}.$$

Since $\sigma_2(J) = 0$, we then have $\pi_1^+ = \pi_1^- = \underline{\pi}_1$. Furthermore, $\underline{\pi}_1 = \tilde{\pi}_1$, where $\tilde{\pi}_1$ (24) is the threshold value of π_1 for capacity growth in the proof of Proposition 3. We conclude that for $\pi_1 > \underline{\pi}_1$, z_+ is decreasing, z_- is increasing, and z_{J^*} is decreasing in q. And for $\pi_1 < \underline{\pi}_1$, z_+ is increasing, z_- is decreasing, and z_{J^*} is increasing in q. ■

Proof of Proposition 5. We prove the result using $z_- = \overline{z}_J - \underline{z}_J$ and $z_+ = \underline{z}_{J+1} - \overline{z}_J$, as defined in expressions (25) and (27) in the proof of Proposition 4. There are two cases, depending on the values of $\sigma_2(i)$.

(i) Suppose $\sigma_2(J) \in (0,1)$ (which implies $\sigma_2(i) > 0$ for $i < J$), or equivalently $q \in \left(\frac{Jh}{1-\pi_2(r_1-\varepsilon)}, \frac{Jh}{1-\pi_2(r_1+\varepsilon)}\right)$. Then $\frac{\partial z_-}{\partial q} > 0$ iff:

$$\pi_1 < \overline{\pi}_1^- \equiv 1 - \frac{h\left[J^2 \frac{\partial \sigma_2(J)}{\partial q} - \sum_{i=1}^{J-1} i \frac{\partial \sigma_2(i)}{\partial q}\right]}{J \frac{\partial \sigma_2(J)}{\partial q} - \sum_{i=1}^{J-1} \frac{\partial \sigma_2(i)}{\partial q}}$$

if the denominator in the preceding expression is positive. To see why the denominator is positive, observe that $\sum_{i=1}^{J-1} \frac{\partial \sigma_2(i)}{\partial q}$ is maximized when all $\sigma_2(i)$ are interior. Thus the denominator is at least:

$$\frac{hJ^2}{2\varepsilon q^2 \pi_2} - \frac{hJ(J-1)}{4\varepsilon q^2 \pi_2},$$

which is clearly positive. By a similar argument, the numerator is positive as well. It is straightforward to verify that $\overline{\pi}_1^- < 1 - Jh$.

Likewise, $\frac{\partial z_+}{\partial q} > 0$ iff:

$$\pi_1 > \overline{\pi}_1^+ \equiv 1 - \frac{h\left[\sum_{i=1}^{J} i \frac{\partial \sigma_2(i)}{\partial q} - (J+1)^2 \frac{\partial \sigma_2(J)}{\partial q}\right]}{\sum_{i=1}^{J} \frac{\partial \sigma_2(i)}{\partial q} - (J+1)\frac{\partial \sigma_2(J)}{\partial q}}.$$

It is straightforward to verify that $\overline{\pi}_1^+ < 1 - (J+1)h$.

We conclude that for $\pi_1 < \hat{\underline{\pi}}_1 \equiv \min\{\overline{\pi}_1^-, \overline{\pi}_1^+\}$, z_+ is decreasing and z_- is increasing in q. Since $\overline{\pi}_1^-, \overline{\pi}_1^+ < 1 - (J+1)h$ and $\tilde{\pi}_1 > 1 - Jh$ (24), z_{J^*} is increasing in q in this case. And for $\pi_1 > \hat{\pi}_1 \equiv \max\{\overline{\pi}_1^-, \overline{\pi}_1^+\}$, z_+ is increasing and z_- is decreasing in q.

(ii) When $\sigma_2(i) = 1$ for all i (for which $\sigma_2(J) = 1$, or equivalently $q \geq \frac{Jh}{1-\pi_2(r_1+\varepsilon)}$, is sufficient), $\frac{\partial z_-}{\partial q} = 0$. Thus there can be no changes in J^* until q increases enough such that $\frac{\partial \sigma_2(J+1)}{\partial q} > 0$. ■

Proof of Proposition 6. Differentiating the expression for $\overline{z}_{J^*}$ (14) with respect to r_1 produces:

$$\frac{\partial \overline{z}_{J^*}}{\partial r_1} = \frac{1}{c(J^*+1)}\left[\sum_{i=1}^{J^*}(1-ih)\left(r_1\frac{\partial \sigma_1(i)}{\partial r_1} + \sigma_1(i)\right) + \sum_{i=1}^{J^*}(1-\pi_1-ih)\left((1-r_1)\frac{\partial \sigma_2(i)}{\partial r_1} - \sigma_2(i)\right)\right].$$

First, we provide sufficient conditions for $\frac{\partial \overline{z}_{J^*}}{\partial r_1} > 0$. Observe that $\frac{\partial \sigma_1(i)}{\partial r_1} > 0$ and $\frac{\partial \sigma_2(i)}{\partial r_1} < 0$ at an interior solution, and $\frac{\partial \sigma_2(i)}{\partial r_1} = 0$ at a corner solution. Furthermore, in equilibrium $1 - J^*h > 0$, and therefore $(1-ih)(r_1\frac{\partial \sigma_1(i)}{\partial r_1} + \sigma_1(i)) > 0$ for all i. Thus, $\sigma_2(i) = 0$ for all i is sufficient for the result. Since $\sigma_k(i)$ is weakly decreasing in i, this condition reduces to $\sigma_2(1) = 0$. Solving for h then produces $h > q(1 - \pi_2(r_1 - \varepsilon))$.

Next, observe that $(1-r_1)\frac{\partial \sigma_2(i)}{\partial r_1} - \sigma_2(i) \leq 0$ for all i. Thus if $1 - \pi_1 < h$ then we also have $\frac{\partial \overline{z}_{J^*}}{\partial r_1} > 0$. Combining the conditions on h yields this part of the result.

Second, to show that $\frac{\partial \overline{z}_{J^*}}{\partial r_1}$ is increasing in π_1 when $\sigma_1(i) = 1$ for all i, we take the partial derivative with respect to π_1:

$$\frac{\partial^2 \overline{z}_{J^*}}{\partial r_1 \partial \pi_1} = \frac{1}{c(J^*+1)} \left[\sum_{i=1}^{J^*} (1-ih) \frac{\partial \sigma_1(i)}{\partial \pi_1} - \sum_{i=1}^{J^*} \left((1-r_1) \frac{\partial \sigma_2(i)}{\partial r_1} - \sigma_2(i) \right) \right]. \tag{29}$$

If $\sigma_1(i) = 1$ for all i, then $\frac{\partial \sigma_1(i)}{\partial \pi_1} = 0$ and by the preceding argument the bracketed term in (29) is positive. Since $\sigma_k(i)$ is weakly decreasing in i, this condition reduces to $\sigma_1(J^*) = 1$ or, solving for π_1, $\pi_1 < \frac{1-hJ^*}{1-(r_1-\varepsilon)}$. ■

9 References

Acemoglu, Daron, Camilo García-Jimeno, and James A. Robinson. 2015. "State Capacity and Economic Development: A Network Approach." *American Economic Review* 105(8): 2364-2409.

Almunia, Miguel, and David Lopez-Rodriguez. 2018. "Under the Radar: The Effects of Monitoring Firms on Tax Compliance." *American Economic Journal: Economic Policy* 10(1): 1-38.

Bachas, Pierre, Roberto Fattal Jaef, and Anders Jensen. 2019. "Size-Dependent Tax Enforcement and Compliance: Global Evidence and Aggregate Implications." *Journal of Development Economics* 140: 203-222.

Berwick, Elissa, and Fotini Christia. 2018. "State Capacity Redux: Integrating Classical and Experimental Contributions to an Enduring Debate." *Annual Review of Political Science* 99(4): 1218-1244.

Besley, Timothy, and Torsten Persson. 2009. "The Origins of State Capacity: Property Rights, Taxation, and Politics." *American Economic Review* 99(4): 1218-1244.

Besley, T and Torsten Persson. 2011. *Pillars of Prosperity: The Political Economics of Development Clusters.* Princeton, NJ: Princeton University Press.

Bolton, Alexander, Rachel A. Potter, and Sharece Thrower. 2016. "Organizational Capacity, Regulatory Review, and the Limits of Political Control." *Journal of Law, Economics, and Organization* 32(2): 242-271.

Brewer, John. 1988. *The Sinews of Power.* Cambridge, MA: Harvard University Press.

Brown, J. David, John S. Earle, and Scott Gehlbach. 2009. "Helping Hand or Grabbing Hand? State Bureaucracy and Privatization Effectiveness." *American Political Science Review* 103(2): 264-283.

Caliendo, Lorenzo, Ferdinando Monte, and Esteban Rossi-Hansberg. 2015. "The Anatomy of French Production Hierarchies." *Journal of Political Economy* 123(4): 809-852.

Cullen, Julie Berry, Nicholas Turner, and Ebonya Washington. 2021. "Political Alignment, Attitudes Toward Government and Tax Evasion." *American Economic Journal: Economic Policy* 13(3): 135-166.

Dal Bó, Ernesto, Frederico Finan, and Martín A. Rossi. 2013. "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service." *Quarterly Journal of Economics* 128(3): 1169-1218.

Derthick, Martha. 1990. *Agency Under Stress.* Washington, DC: The Brookings Institution.

Dincecco, Mark, and Gabriel Katz. 2014. "State Capacity and Long-Run Economic Performance." *Economic Journal* 126(590): 189-218.

Finan, Frederico, Benjamin A. Olken, and Rohini Pande. 2017. "The Personnel Economics of the Developing State." In ed. Abhijit Banerjee and Esther Duflo, *Handbook of Economic Field Experiments* Vol. 2, 467-514.

Foarta, Dana. 2021. "How Organizational Capacity Can Improve Electoral Accountability." Unpublished manuscript, Stanford University.

Fukuyama, Francis. 2013. "What Is Governance?" *Governance* 26(3): 347-368.

Gailmard, Sean, and John W. Patty. 2012. "Formal Models of Bureaucratic Politics." *Annual Review of Political Science* 15: 353-377.

Garfias, Francisco. 2018. "Elite Competition and State Capacity Development: Theory and Evidence from Post-Revolutionary Mexico." *American Political Science Review* 112(2): 339-357.

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy* 108(5): 874-904.

Garicano, Luis, and Esteban Rossi-Hansberg. 2006. "Organization and Inequality in a Knowledge Economy." *Quarterly Journal of Economics* 1383-1435.

Garicano, Luis, and Esteban Rossi-Hansberg. 2012. "Organizing Growth." *Journal of Economic Theory* 147: 623-656.

Gennaioli, Nicola, and Hans-Joachim Voth. 2015. "State Capacity and Military Conflict." *Review of Economic Studies* 82(4): 1409-1448.

Gibbons, Robert, and John Roberts. 2014. "Economic Theories of Incentives in Organizations." In ed. R. Gibbons and J. Roberts, *The Handbook of Organizational Economics.* Princeton, NJ: Princeton University Press.

Grindle, Merilee. 2012. *Jobs for the Boys: Patronage and the State in Comparative Perspective.* Cambridge, MA: Harvard University Press.

Grossman, Matthew, and David A. Hopkins. 2016. it Asymmetric Politics: Ideological Republicans and Group Interest Democrats. New York: Oxford University Press.

Huber, John D., and Nolan M. McCarty. 2004. "Bureaucratic Capacity, Delegation, and Political Reform." *American Political Science Review* 98(3): 481-494.

Johnson, Hubert C. 1975. *Frederick the Great and his Officials.* New Haven, CT: Yale University Press.

Johnson, Noel D., and Mark Koyama. 2017. "States and Economic Growth: Capacity and Constraints." *Explorations in Economic History* 64: 1-20.

Kasara, Kimuli, and Pavithra Suryanarayan. 2020. "Bureaucratic Capacity and Class Voting: Evidence from Across the World and the United States." *Journal of Politics* 82(3): 1097-1112.

Mann, Thomas E. and Norman J. Ornstein. 2012. *It's Even Worse than It Looks.* Basic Books, 2012.

Moe, Terry M. 1989. "The Politics of Bureaucratic Structure." In ed. John E. Chubb and Paul E. Peterson, *Can the Government Govern?* Washington, DC: The Brookings Institution.

Propublica. 2018. "After Budget Cuts, the IRS' Work Against Tax Cheats Is Facing 'Collapse'," October 1, 2018. (https://www.propublica.org/article/after-budget-cuts-the-irs-work-against-tax-cheats-is-facing-collapse).

Propublica. 2020. "Has the IRS Hit Bottom?" June 30, 2020. (https://www.propublica.org/article/has-the-irs-hit-bottom).

Schreiber, Leon. 2019. "Bolstering Revenue, Building Fairness: Uganda Extends its Tax Reach, 2014-2018." Innovations for Successful Societies, Princeton University (http://successfulsocieties.princeton.edu).

Silberman, Bernard S. 1993. *Cages of Reason: The Rise of the Rational State in France, Japan, the United States, and Great Britain.* Chicago: University of Chicago Press.

Thijs, Nick, Gerhard Hammerschmid, and Enora Palaric, 2017. "A Comparative Overview of Public Administration Characteristics and Performance in EU28." European Commission Directorate-General for Employment, Social Affairs and Inclusion (https://ec.europa.eu/social/BlobServlet?docId=19208).

Tilly, Charles. 1990. *Coercion, Capital, and European States, AD 990-1990.* Cambridge, MA: Blackwell.

Ting, Michael M. 2011. "Organizational Capacity." *Journal of Law, Economics, and Organization* 27(2): 245-271.

United States Congressional Budget Office. 2020. "Trends in the Internal Revenue Service's Funding and Enforcement." (https://www.cbo.gov/publication/56467).