American Political Science Review (2021) 115, 2, 667–685
doi:10.1017/S0003055421000046

The Political Economy of Governance Quality

MICHAEL M. TING *Columbia University*

This paper develops a dynamic theory of the social and political foundations of governance quality. In the model, groups of citizens have different expected needs for a public service, and citizens choose whether to demand service when the need arises. Politicians representing these groups can determine policy benefits and delegate to bureaucrats the ability to invest in long-run service quality. The main feature of the theory is its foundation for citizen–government interactions, which draws from well-known queueing models of organizational service provision. The model provides a framework for characterizing the effectiveness and durability of government programs. A main implication is that politicized bureaucracies improve program survivability and increase the frequency of investment, while insulated bureaucracies increase the intensity of investment; overall service quality trades off between these two factors. Other results examine the implications of cross-group inequality, electoral conditions, and decentralization.

INTRODUCTION

Modern governments produce a vast array of essential services. Even the partial list of permits, public health measures, education, transportation, and law enforcement evinces the everyday centrality of the state in any functioning society. Good governance links needed outputs with eligible citizens, for citizens cannot benefit from inaccessible services. The 2020 pandemic starkly exhibited its importance, as communities in the United States and elsewhere quickly felt the consequences of successes and failures in providing financial assistance, medical tests, and even safe voting procedures.

What are the social and political bases of governance quality? The set of possible contributing factors is extensive, ranging from local citizen oversight to national political systems (e.g., Pepinsky, Pierskalla, and Sacks 2017). This paper develops a general theory of governance quality that takes access to state services as a starting point and then considers how political and institutional factors shape provision over time. Its primary metric of quality is waiting times, which have the virtues of pervasiveness and empirical observability. Government agencies use wait times to measure performance in areas as diverse as health care, airport security, voting, and disability benefits (e.g., Government Accountability Office 2014; 2015; 2018; Social Security Administration 2018), and many academic analyses have followed suit (e.g., Ando 1999; Bolton, Potter, and Thrower 2016; Carpenter 2002; 2004; Whitford 2005).

Examples of consequential wait times are easy to find. In some countries, lengthy lines are a defining feature of the relationship between citizens and the state. In 2018 the Inter-American Development Bank (IDB) reported that Latin American residents spent an average of over five hours per government transaction, with some transactions requiring multiple trips over several days (IDB 2018).[1] Bad service provision is more than an inconvenience. As the IDB report notes, wait times are both regressive and corrupting, often deterring low-income citizens from completing transactions while giving incentives for wealthier citizens to bribe bureaucrats.

Two U.S. examples further illustrate how waiting times affect even the most important public policies. First, the Social Security Administration (SSA), which manages the country's popular social insurance and disability programs, lost 11% of its inflation-adjusted operating budget between 2010 and 2017. Notably, these cuts did not affect citizens' statutory benefit levels, even as demographic trends increased the agency's caseload by 15%. The resulting hiring freezes, reduced overtime, and understaffed call centers contributed to a disability hearing backlog of over 1.1 million people, with an anticipated wait duration of 21 months (Romig 2017). Second, some urban police departments face chronic difficulties in responding rapidly to 911 emergency calls. These delays grew in salience following the 2008 recession, which caused staff reductions in cities such as New Orleans and Detroit. Slow police responses have contributed to failures in recording and solving cases (Asher 2018; Blanes i Vidal and Kirchmaier 2018).

Managing service demand is hardly unique to the public sector; firms and other organizations face similar challenges in servicing clients or customers. Since the mid-twentieth century, analysts have employed formal models of *queues* to study service center operations. In the simplest queueing model, clients arrive at a provider randomly over time and depart once served. If the provider is busy servicing a previous client, then a queue forms and clients must take the anticipated cost

Michael M. Ting, Professor of Political Science and International and Public Affairs, Department of Political Science and SIPA, Columbia University, mmt2033@columbia.edu

Received: July 02, 2019; revised: December 14, 2020; accepted: January 15, 2021. First published online: March 8, 2021.

[1] The transactions concerned areas including identification, registration, education, health, tax payment, and pensions. The IDB report synthesized data from the 2017 Latinobarómetro survey, which asked about wait times for the first time that year.

of waiting into account. The rate of arrivals depends on the characteristics of the population, and the rate of departures depends on the capacity of the provider. High-capacity providers resolve cases in less time, thereby reducing congestion and improving client welfare. By capturing a commonly experienced aspect of service quality, queueing models have become standard tools for analyzing the performance of systems ranging from customer service desks to computer networks (Gross et al. 2008).

Despite their intuitive elegance and analytical tractability, existing queueing models are generally unsuitable for describing political settings. While virtually all citizens are potential users of public services, the analogy with customer service breaks down in two important ways. First, political environments typically feature heterogeneity in demand: some groups want more service, others less. Some citizens or their representatives might therefore be unwilling to use government resources to fulfill the needs of others. Second, the political systems that decide the level and accessibility of services generate distinct incentives unlike those in firms or other organizations. Thus even seemingly obvious reforms such as the IDB (2018) recommendation of increasing electronic transactions across Latin America must gain the assent of actors with different objectives and time horizons.

To adapt queueing models for government services, this paper embeds queues in a simple dynamic policymaking model. Its framework incorporates three well-recognized aspects of the relationship between politicians and bureaucrats. The first is bureaucratic appointment structures (e.g., Moe 1989). Some bureaucracies are politicized, in the sense of having leaders who serve while their appointing politicians hold office. Others are politically insulated and have leaders who serve regardless of election results. These matter because the prospect of electoral turnover affects time horizons and may reduce the appeal of long-term service improvements. Modern bureaucracies in wealthy democracies are hybrids of these systems, and considerable variation exists across U.S. federal and state agencies. The second is bureaucratic investment. Bureaucrats are often uniquely knowledgeable about the technological and operational potential of their agencies, and thus their effort is critical for achieving service improvements (e.g., Rosen 1988). The final feature is delegation, which is perhaps the canonical tool for controlling the bureaucracy (e.g., Epstein and O'Halloran 1999). Politicians often cannot simply mandate better service, but they can choose whether to allow bureaucrats to invest in costly improvements.

In the model, two groups of citizens live in continuous time, divided into periods of length one. At random times, individual citizens become eligible for a public service, such as a license. To capture heterogeneous demands, arrival rates are identical within groups but different across groups. An affected citizen chooses immediately whether to queue before a government bureaucrat. If she queues, then she must wait at a cost for the bureaucrat to resolve all preceding cases as well as her own. As with the familiar experiences of interacting with call centers or customer service departments, citizens do not observe the length of the queue. Thus they must form rational conjectures about whether joining is beneficial.

In each period, a politician representing one of the groups wins election with some exogenous probability. Politicians can either provide benefits for eligible citizens and basic administrative support (e.g., payroll) or shut down the program.[2] They also face dynamic incentives regarding service quality. By delegating, the politician gives the bureaucrat authority to make capacity investments that improve service in the subsequent period. Investments impose costs on both the bureaucrat and the delegating politician. By not delegating, the politician foregoes investment and capacity depreciates. Benefits, administration, and investment are all financed by a flat tax on the two groups. Politicians can run for reelection once and maximize the welfare of their constituents during their time in office.

Bureaucrats are public-service motivated and invest in order to reduce citizen waiting times. Each bureaucrat can live for up to two periods, which allows the model to capture both natural cycles of organizational aging and different appointment rules. An insulated bureaucrat always serves two periods, independently of election results, while a politicized bureaucrat enters office with each first-term politician and departs if she loses reelection. The former might represent an agency with strong civil service protections or a commission-like structure, while the latter might correspond to an agency with more political appointees or other avenues for political influence.

In equilibrium, a politician will provide service if her group has sufficient demand for it and capacity is high enough. Her decision over delegation is more complex. She will typically delegate only if capacity is *intermediate*—low enough to require strengthening, but high enough to make rescuing the program affordable. A politician may even delegate while the program is shut down, in anticipation of a future restart. Crucially, delegation also depends on age or whether actors are in the first or second period of their careers. Since investment can only produce future benefits, second-term politicians have no incentive to delegate and second-term bureaucrats have no incentive to invest. Without sustained investments, capacity may depreciate to the point where neither group would want to revive the program, at which point it effectively ends.

The need for high and frequent investment produces the model's central tension. As Rauch (1995) and Gailmard and Patty (2007) observed, insulated programs receive higher investments, since the bureaucrat is certain of staying in office to enjoy its benefits. By contrast, politicized programs receive more frequent

[2] The Oklahoma attorney general's office provides an example of how politicians adapt services to constituent interests. In 1996, Democratic Attorney General Drew Edmondson created an Environmental Protection Unit to investigate and litigate environmental law violations. In 2012, his Republican successor Scott Pruitt effectively ended its activities.

investments, since every newly elected politician freshly appoints an aligned bureaucrat. This trade-off reconciles two long-standing perspectives on the effects of bureaucratic personnel. In various contexts, studies have demonstrated that civil servants—a key means of insulation—improve bureaucratic performance (e.g., Carpenter 2001; Lewis 2007). However, civil servants do not always outperform political appointees (Aberbach and Rockman 2000; Krause, Lewis, and Douglas 2006), and even highly insulated agencies face some political influence (Moe 1982). Intensive exposure to political direction can generate ideas and improve monitoring (Bilmes and Neal 2003; Maranto 1998; Moe 1985; Raffler 2019). Political appointees therefore often provide the impetus for major policies; for example, President Obama appointed Jeffrey Zients to lead efforts to fix his 2009 "Cash for Clunkers" auto efficiency initiative and the troubled 2013 rollout of the Obamacare web portal.[3] More broadly, Hollibaugh, Horton, and Lewis (2014) show that the Obama administration placed its most competent political appointees in agencies it prioritized, and Rogger (2018) shows that greater political oversight generated higher rates of project implementation in Nigerian bureaucracies.

The investment dynamics interact with political and social conditions to produce governance quality. A primary question is whether programs can survive at all in the long run. Conventional wisdom once held that government organizations were nearly immortal (e.g., Downs 1967; Kaufman 1976), but more recent studies have identified partisan shifts and fiscal deficits as causes of termination (Berry, Burden, and Howell 2010; Carpenter and Lewis 2004). This model adds two perhaps counterintuitive predictions. First, inequality in demand across groups *helps* survival, as it concentrates supporters into a group whose politicians are inclined toward investment. An electoral advantage for this group further enhances survivability. Second, insulation can hinder survivability, as matches between first-term politicians and young bureaucrats may be too infrequent to ensure sufficient investment.

Conditional upon long-run survival, the Markov process governing the evolution of capacity generates novel predictions about how political insulation and the electoral environment affect long-run quality. A competitive electoral environment maximizes the quality of insulated programs, but politicized programs might perform better when elections are less competitive. In a competitive electorate, insulated programs perform better than politicized ones when capacity is durable (i.e., depreciation is slow), but this advantage reverses when capacity is less durable. Thus, neither political accountability nor insulation from it unambiguously benefit citizens.

The model provides a flexible framework for examining service quality in a variety of institutional settings. A numerical extension illustrates this by examining an alternative setting with decentralized, group-specific programs. Decentralization can increase average quality because each group can tailor the program to its own needs. However, decentralized groups may sometimes be unable to sustain programs on their own. Centralization can boost investment by allowing the group in power to pass investment costs onto the opposition. This can ensure program survival when decentralized programs would be unviable.

The paper proceeds as follows. The following subsection reviews related literature. The next section describes queues along with the model, and the subsequent section derives queueing, investment, delegation, and policy strategies. The penultimate section combines these components to derive results on long-run program survival and quality. The final section proposes avenues for further inquiry and concludes.

RELATED LITERATURE

This paper examines the foundations of governance quality by integrating citizens, bureaucrats, and politicians as strategic actors. To a significant degree, existing work has focused on either relationships between citizens and the state or between politicians and bureaucrats. The former relationship starts with direct contact between citizens and bureaucrats. Lipsky (1980) brought "street-level bureaucracy" into the disciplinary lexicon and observed how on-the-ground constraints could create substantial frictions in the implementation of public policies. Herd and Moynihan (2018) show how administrative burdens on citizens produce indirect political and distributive consequences. A burgeoning new generation of field studies has shed light on how mechanisms such as monitoring, information, and technology affect service provision (e.g., Bussell 2019; Kruks-Wisner 2018; Slough 2020; Wilke 2020). The broader electoral context also matters; Keefer and Khemani (2005) review how factors such as polarization and asymmetric information distort service provision.

Despite the common usage of waiting times in measuring bureaucratic outputs, the theoretical literature on bureaucracy-client relationships has focused predominantly on the problem of determining client eligibility under asymmetric information (e.g., Banerjee 1997; Prendergast 2003; Ting 2017). As argued above, queueing is a natural alternative approach to service provision, and researchers have applied it to a wide variety of organizations. Examples with possible policy implications include tolls (Naor 1969), bribery (Lui 1985), and hierarchies (Beggs 2001). However, queueing models generally do not consider political environments, and few applications exist in political science. Two papers that do invoke explicit institutional settings but do not develop theoretical models are Ando (1999), who studies endangered species classification, and Herron and Smith (2016), who study voting administration.

The political control of bureaucracies is the subject of an extensive family of agency models (Gailmard and Patty 2012). Within it, two themes are particularly

[3] In 2020, President-elect Biden appointed Zients as head of his COVID-19 task force.

relevant. First, several papers focus on the role of capacity or valence as distinct from policy outcomes (Besley and Persson 2009; Hirsch and Shotts 2012; Huber and McCarty 2004; Turner 2019). Second, two recent papers consider the evolution of policy quality over time. Callander and Martin (2017) model a setting with policy decay, while Gratton et al. (2020) explore electoral incentives for legislating when bureaucratic quality affects the quality of laws.

The model draws some of its main features from existing theoretical and empirical accounts of bureaucratic behavior. Perhaps the most important of these is political insulation. Among other effects, empirical studies have found that insulating personnel through civil service regulations increases investment (Rauch 1995), reduces corruption (Dahlström, Lapuente, and Teorell 2012), and improves measures of program performance (e.g., Lewis 2007), but it also blunts the initiative of elected or appointed officials (Heclo 1977). Insulation becomes particularly important in conjunction with elections, which affect politicized bureaucrats' time horizons and thus their incentives to perform. A series of papers examines the interaction between elections and civil service reform (Horn 1995; Mueller 2015; Ting et al. 2013; Ujhelyi 2014), while Nath (2015) and Akhtari, Moreira, and Trucco (2020) provide evidence that political turnover hurts bureaucratic output.

Other shared features include investment, delegation, and public-service motivation. Bureaucratic investment flows from expertise: many studies have demonstrated the bureaucracy's role as the critical investor in policy technologies (e.g., Carpenter 2001; Rosen 1988). By contrast, the inability to harness bureaucratic capabilities can impede politicians' policy objectives (e.g., Bolton, Potter, and Thrower 2016; Derthick 1990). Models of delegation (e.g., Epstein and O'Halloran 1999) typically consider the trade-off between expertise and ideological affinity. Here, however, the bureaucrat is an expert but her conflict with politicians arises from costs and time horizons rather than ideological conflict. Finally, public-service motivation is a common explanation for bureaucratic productivity in the absence of ideological motivations or strong contractual incentives (Dal Bó, Finan, and Rossi 2013; Francois 2000; Rainey and Steinbauer 1999).

MODEL

The model incorporates service provision, investment, and elections over continuous time, divided into periods of duration 1. Where necessary, periods are denoted with a subscript. The interaction between the bureaucracy and citizens is modeled as a queueing process, whereby citizens who qualify for a public service can join a queue in order to receive it. I first describe the queueing process, and then the political environment.

Queueing for Service

A basic queueing model describes a service organization. It consists of an *arrival process* that generates citizen demands and a *solution process* whereby the organization resolves them.

The arrival process works as follows. There are two groups in society, labeled 1 and 2, each populated by a continuum of measure 1 of citizens. Group i citizens become eligible for service according to a Poisson process with rate λ_i, where $\lambda_1 < \lambda_2$. I interpret the different rates as partisan inequality and call groups 1 and 2 the low- and high-demand groups, respectively. As examples, SSA disability caseloads are higher in older communities, while police response times are of greater concern in high-crime communities. By the well-known properties of the Poisson distribution, group i produces λ_i cases per period in expectation, with a realized number of cases X_i distributed according to

$$\Pr\{X_i = n\} = \frac{\lambda_i^n}{n!} e^{-\lambda_i}.$$

By the additive property of the Poisson distribution, the aggregate arrival rate of cases in the population is $\Lambda = \sum_i \lambda_i$. Using standard formulas, the expected time interval between cases is exponentially distributed, with density $\Lambda e^{-\Lambda\tau}$ and mean $1/\Lambda$. Since each period has duration 1, with probability $e^{-\Lambda}$ there are no arrivals in a given period. Thus for λ_i sufficiently large, the probability of no arrivals is negligible.

The onset of a case makes a citizen eligible for the public service. Citizens choose whether to join the queue, but observe neither its length nor the actions of other citizens. Joining is irreversible, and thus a citizen must stay in line until her case is resolved. A citizen who waits a total duration τ for service (from waiting for both her case and previously queued cases to finish) experiences a cost of $c\tau$, where $c \geq 0$. Citizens are risk neutral and receive a payoff of $b_t \geq 0$ from resolution, where b_t is determined by the incumbent politician. Benefits and waiting costs are identical for all citizens.[4] Not joining results in a payoff of zero.

The solution process corresponds to a bureaucracy that resolves queued cases in a first-come first-serve (FCFS) manner. There is only one servicer or bureaucrat, so each queued case must wait for the completion of all preceding cases from that period. Solutions follow a Poisson process, with rate $\mu_t > 0$. Analogously to the arrival times, expected service times are exponentially distributed, with density $\mu_t e^{-\mu_t \tau}$ and mean $1/\mu_t$. The parameter μ_t represents the organization's technology or *capacity* in period t. The bureaucrat resolves all queued cases that originate within period t according to μ_t, even if solution times spill over the period's duration of 1.

Together, these components define a FCFS $M/M/1$ (for Markov arrival, Markov solution, one server)

[4] The assumptions of identical costs and benefits across groups and unobservable queue length simplify the analysis by giving all citizens the same incentives, conditional upon having a need for service. Heterogeneity in relative benefits may result in only one group joining the queue. Imposing a fixed cost on citizens for each case would not affect the results.

queue, which is commonly regarded as the most elementary queueing process. The limiting properties of this Markov process are both simple and standard, and they ensure that long-run behavior is independent of the current status of the queue.[5] Under the assumption that all arrivals join the queue, several of the most important properties are as follows:

- Utilization (the proportion of time spent servicing clients): $\rho = \frac{\Lambda}{\mu_t}$
- Average number of customers in the queue and in service: $\frac{\rho}{1-\rho} = \frac{\Lambda}{\mu_t - \Lambda}$
- Probability of having n clients in the queue: $p_n = \lim_{t \to \infty} \Pr\{X(t) = n\} = (1-\rho)\rho^n = \rho^n p_0$
- Average waiting time upon joining a queue:

$$W(\mu_t) = \frac{1}{\mu_t - \Lambda}. \tag{1}$$

Observe that unless $\mu_t > \Lambda$, the size of the queue grows without limit, and thus an effective service organization must satisfy this constraint.[6]

Political Process

An infinite horizon political process determines the features of the service queue. In each period, a politician from one of the two groups takes office. The probability of election for group i is exogenously fixed at $\pi_i \in (0, 1)$, with $\pi_1 = 1 - \pi_2$. This exogeneity captures the assumption that other issues overshadow bureaucratic performance in determining election outcomes. Each politician lives for up to two periods and stands for reelection immediately after her first term.

A politician begins period t by choosing whether to offer the public service $s_t \in \{0, 1\}$, the benefit level $b_t \geq 0$ that citizens receive from resolved cases, and whether to delegate investment authority $d_t \in \{0, 1\}$ to the bureaucrat. It is useful to consider s_t and b_t as the period t *policy.* Offering the service ($s_t = 1$) provides the bureaucracy with short-term administrative resources such as payroll, consumable supplies, and overhead that allow it to distribute benefits b_t.

Delegation affects the bureaucracy's problem-solving ability. As described earlier, capacity μ_t determines the queue's solution rate. Initial capacity μ_1 might reflect factors such as the quality of the government's personnel. If (and only if) she is delegated authority ($d_t = 1$), the bureaucrat makes an investment choice $e_t \geq 0$. Investment adds to capacity, but capacity depreciates over time: a proportion $\delta \in (0, 1]$ survives into the next period. Thus δ is a measure of the durability of the program's personnel or physical capital. Capacity evolves according to

$$\mu_{t+1} = \delta(\mu_t + e_t d_t). \tag{2}$$

The public budget includes the cost of administration, benefits, and delegated spending. When a politician provides the service ($s_t = 1$), its cost depends on both the arrival rate of queued cases and the probability with which citizens queue for service. If each eligible citizen joins the queue with probability q, then the effective arrival rate is $q\Lambda$. Each queued case imposes a fixed administrative cost $k > 0$ and a direct cost b_t^2 of benefit provision. The increasing marginal cost of benefits arises from greater opportunities for mismanagement or fraud, and hence higher monitoring expenses.[7] These opportunities are of special concern for high-value services such as tax exemptions. As the politician provides the bureaucrat's resources, the budget also includes a portion $\kappa_p \in (0, 1)$ of the bureaucrat's costs $e_t d_t$. The politician's total expected period t budget is

$$\kappa_p e_t d_t + q\Lambda(k + b_t^2).$$

All expenditures are covered by a tax that is distributed evenly between groups.

When a politician does not offer the service ($s_t = 0$), the program shuts down and there are no administrative or benefit costs. However she may still delegate investment authority, which could increase capacity for period $t + 1$.

Politicians care about the welfare of their respective groups over the periods during which they hold office. Since only group-level welfare matters in the model, there is no need to specify the distribution of taxes within groups. When $s_t = 1$ and all eligible citizens willingly queue, the expected welfare of group i in period t is

$$u_i(b_t, d_t; e_t, \mu_t) = \lambda_i \left(b_t - \frac{c}{\mu_t - \Lambda} \right) - \frac{\kappa_p e_t d_t + \Lambda(k + b_t^2)}{2}. \tag{3}$$

The model explores two kinds of program leadership structures. In one, bureaucrats are like civil servants and stay in office for two periods before retiring. Thus their time horizons are independent of election results. Alternatively, bureaucrats are like political appointees whose term of office coincides with those of politicians. Thus a bureaucrat reaches age 2 if and only if her appointing politician wins reelection. I refer to the former as *insulated*, and the latter as *politicized.* Since the bureaucrat must be in office to benefit from an investment, it will be convenient to adopt the following notation for her probability of reaching age 2.

$$\pi_b = \begin{cases} \pi_i & \text{if politicized, group } i \text{ politician} \\ 1 & \text{if insulated.} \end{cases} \tag{4}$$

[5] Using the limit properties is standard in queueing models. For any finite interval of time, the limit properties approximate the parameters of the queue.

[6] A queue with a capacity (length) constraint does not require $\mu_t > \Lambda$, since any arrivals when a queue is at capacity are not served.

[7] Any convex cost function would produce similar results, while concave costs would result in corner solutions for benefits.

The bureaucrat is public-service motivated and cares about client waiting times, which directly affects client welfare and may indirectly affect perceived organizational competence. She only has a strategic choice if the politician delegates, and makes no decisions affecting current clients. Her payoff in a single period is

$$u_b(e_t;\mu_t) = -\frac{m_b}{\mu_t - \Lambda} - \kappa_b e_t, \tag{5}$$

where $\kappa_b > 0$ is the bureaucrat's marginal cost of effort and $m_b \in [\lambda_1, \lambda_2]$ is a measure of her public-service motivation. Although not examined in this paper, variations in m_b can potentially reflect group-based bureaucratic preferences.

I impose two parametric assumptions that simplify the analysis by reducing the number of corner solutions. First, the following condition ensures that the bureaucrat's optimal investment will induce politicians of both groups to provide service.[8]

$$\sqrt{\frac{\delta \pi_b m_b}{\kappa_b}} > \frac{2c\lambda_1 \Lambda}{\lambda_1^2 - k\Lambda^2}. \tag{6}$$

High public-service motivation and low investment costs ensure this condition. Second, to prevent shutting down the program from being dominant, let $k < (\lambda_1/\Lambda)^2$. Note that this condition implies that the right-hand side of Equation 6 is positive.

All actions in the game are observable, aside from current-period queueing choices. Table 1 lists the main parameters of the model. In each period, the sequence of moves is as follows:

1. Nature elects or reelects the group i politician with probability π_i.
2. Nature appoints or reappoints a bureaucrat according to the personnel selection rule.
3. The politician chooses program status s_t, policy benefit b_t, and delegation d_t.
4. If delegated authority, the bureaucrat chooses investment e_t.
5. Nature draws eligible citizens according to rates λ_1 and λ_2; eligible citizens choose immediately whether to queue.

I characterize a subgame perfect equilibrium that is symmetric in citizen queueing strategies. Let H_t denote the history of all actions through the period t election. In period t, the politician's strategy is a mapping $H_t \to \{0,1\}^2 \times \mathbb{R}_+$ into choices of s_t, d_t, and b_t, respectively. The bureaucrat's investment strategy is a mapping $H_t \times \{0,1\} \times \mathbb{R}_+ \to \mathbb{R}_+$ into an investment e_t. Finally strategies for citizens are mappings $H_t \times \{0,1\}^2 \times \mathbb{R}_+^2 \to [0,1]$ into a probability of joining the queue if a case arises.

[8] Since $\lambda_1 \leq \Lambda/2$, (6) can be satisfied only if $\sqrt{\frac{\delta m_b}{\kappa_b}} > \frac{4c}{1-4k}$.

TABLE 1. Notation

λ_i	group i demand rate
Λ	total social demand rate
μ_t	period t solution rate (capacity)
π_i	group i election probability
π_b	probability of a first-term bureaucrat staying in office
b_t	period t policy benefit
s_t	period t service decision
d_t	period t delegation
e_t	period t bureaucratic investment
m_b	bureaucrat's public-service motivation
κ_b	bureaucrat's marginal cost of investment
κ_p	politician's marginal cost of investment
k	politician's per-unit cost of provision
c	citizen's marginal waiting cost
δ	durability

RESULTS

This section first examines a single period of the game, which identifies citizens' queueing incentives and politicians' policy incentives. It then addresses the infinite horizon model, which produces dynamic incentives through delegation and investment.

One Period

The single-period setting completely describes periods with reelected politicians, who are unconcerned with future service capacity. For notational convenience I suppress time subscripts for this subsection.

The first step is to characterize strategies of eligible citizens. Working backward, queueing citizens receive a policy benefit b but also face waiting costs, which depend on capacity μ and the behavior of other citizens. Suppose that each eligible citizen independently joins the queue with probability q (where q may be 0 or 1). Since queued citizens must stay in line until their cases and those of all predecessors are resolved, an eligible citizen is indifferent between joining and not joining if the benefit equals the cost of her waiting time, as given by Equation 1:

$$b = \frac{c}{\mu - q\Lambda}, \tag{7}$$

where $q\Lambda$ is the effective population demand rate.

Benefits that exceed the threshold in Equation 7 would cause a citizen to join with certainty, while lower benefits would cause her to avoid the queue. Solving for q produces a unique symmetric equilibrium queueing probability for eligible citizens:

$$q^* = \begin{cases} 0 & \text{if } b < \dfrac{c}{\mu} \\ \dfrac{\mu - c/b}{\Lambda} & \text{if } b \in \left[\dfrac{c}{\mu}, \dfrac{c}{\mu - \Lambda}\right) \\ 1 & \text{if } b \geq \dfrac{c}{\mu - \Lambda}. \end{cases} \tag{8}$$

Now consider the politician's problem. A preliminary question is whether to offer the program at all. Since providing service imposes fixed costs, the politician is better off shutting down the program ($s = 0$) if citizens do not strictly gain from queueing. Effective service provision therefore requires benefits and capacity to be large enough to induce queueing with certainty, or $q^* = 1$.

The politician cannot add to the current period's capacity, and thus there is neither delegation ($d^* = 0$) nor investment ($e^* = 0$). This leaves benefits as her sole lever. To determine this, it is easily verified that her objective as presented in Equation 3 when all citizens queue is concave. Taking the first-order condition and solving for b produces

$$b_i^* = \frac{\lambda_i}{\Lambda}. \tag{9}$$

The optimal benefit depends only on the politician's favored group's relative demand for the public service, not on capacity.

The politician then provides service if b_i^* is generous enough to produce higher utility than shutting down. This will be true under the following condition on existing capacity:

Definition 1. *A program is viable if*

$$\mu > \underline{\mu}_i \equiv \Lambda + \frac{2c\lambda_i\Lambda}{\lambda_i^2 - k\Lambda^2}. \tag{10}$$

The right-hand side of Equation 10 is the capacity threshold for shutting down a program. Viability becomes easier to satisfy as the incumbent politician's demand increases relative to that of the opposition and as waiting costs (c) decrease.

The first result combines these derivations to characterize government outputs in a single period. Politicians open programs if they are viable, and set benefits that are proportional to their constituents' demands. All proofs can be found in the appendices.

Proposition 1. Policies in a Single Period. *There is no delegation ($d^* = 0$). Under politician i, policies are*

$$\left(s_i^*, b_i^*\right) = \begin{cases} (0, 0) & \text{if } \mu \leq \underline{\mu}_i \\ \left(1, \dfrac{\lambda_i}{\Lambda}\right) & \text{if } \mu > \underline{\mu}_i. \end{cases} \tag{11}$$

The politician's expected utility from a single period without delegation is

$$u_i\left(b_i^*, 0; 0, \mu\right) = \begin{cases} 0 & \text{if } \mu \leq \underline{\mu}_i \\ \dfrac{\lambda_i^2}{2\Lambda} - \dfrac{c\lambda_i}{\mu - \Lambda} - \dfrac{k\Lambda}{2} & \text{if } \mu > \underline{\mu}_i. \end{cases} \tag{12}$$

The proof of the result shows that the constraint of inducing queueing with probability one is not binding, as taxation and administrative costs cause the politician to receive less than her reservation utility if citizens use mixed strategies. Thus she would shut down a program before the point at which citizens become indifferent between queueing and staying home. Doing so is by definition ex ante optimal for her group, but not necessarily for the opposing group or society as a whole.

Reelection, Delegation, and Investment

Unlike second-term politicians, newly elected politicians may benefit from bureaucratic investment. Their delegation decisions will drive long-term service quality under alternative political conditions and appointment structures.

An important initial observation is that the queueing and policy choices from the previous subsection are identical under first-term politicians. For citizens, queueing decisions affect neither the election nor the bureaucracy's capacity. This frees them to maximize their short-run payoffs when they are eligible to queue, based on benefits and expected waiting costs in the current period.

A first-term politician's strategy anticipates her possible reelection. Substituting in her second-term strategy from Proposition 1, her objective is

$$\begin{aligned} U_i(b_t, d_t; e_t, \mu_t) = {} & u_i(b_t, d_t; e_t, \mu_t) \\ & + \pi_i u_i\left(b_i^*, 0; 0, \delta(\mu_t + e_t d_t)\right), \end{aligned} \tag{13}$$

where b_i^* is the single-period solution given by Equation 11. It is clear from this equation that the period t opening and benefit choices (s_t and b_t) cannot affect her period $t + 1$ payoffs. Importantly, despite the cost of bureaucratic investments, delegation also does not affect current policy choices. This is because the additive separability of investment costs (seen in Equation 3) makes the marginal return of providing service independent of investment. Thus the politician chooses policies myopically in every period, as summarized in Proposition 1.

This leaves delegation and investment as the remaining choices to characterize. Working backward, it is clear that just as reelected politicians have no incentive to delegate, age-2 bureaucrats have no incentive to invest ($e_t^* = 0$). Nontrivial investment and delegation can therefore occur only when newly elected politicians face age-1 bureaucrats.

Under delegation ($d_t = 1$), an age-1 bureaucrat's general objective for both insulated and politicized programs is

$$U_b(e_t; 1, \mu_t) = u_b(e_t; \mu_t) + \pi_b u_b(0; \delta(\mu_t + e_t)). \tag{14}$$

This objective is concave and produces a straightforward investment solution. The optimal investment raises capacity to the following threshold value, as long as existing capacity starts below it:

$$\mu_b^0(\pi_b) = \frac{\Lambda}{\delta} + \sqrt{\frac{\pi_b m_b}{\delta \kappa_b}}. \tag{15}$$

When existing capacity exceeds this threshold, further investment would not benefit the bureaucrat and the politician would obviously gain nothing from delegation.

The next result summarizes the preceding optimal policy and investment choices.

Lemma 1. Policy and Investment Under Delegation. *Politician i's policy choices in each period t are given by Proposition 1. If she delegates, bureaucratic investment is*

$$e_t^* = \begin{cases} \mu_b^0(\pi_b) - \mu_t & \text{if } \mu_t < \mu_b^0(\pi_b) \text{ and bureaucrat is age 1} \\ 0 & \text{otherwise.} \end{cases} \quad (16)$$

Lemma 1 implies that investment is politically sensitive. Whether from insulation or affiliation with an electorally favored party, bureaucrats with a higher probability of remaining in office (π_b) will invest more.[9] Despite some significant differences in the policy setting, this motivation for investment resembles that of Gailmard and Patty (2007). Investments are also increasing in the bureaucrat's public-service motivation (m_b).

Following investment, capacity depreciates and applying Equation 2 produces the next period's capacity level:

$$\mu_{t+1} = \delta\mu_b^0(\pi_b). \quad (17)$$

Assumption in Equation 6 ensures that the program is viable at the updated capacity level (i.e., $\delta\mu_b^0(\pi_b) > \underline{\mu}_i$).

With the implications of investment established, the final decision is whether to delegate. Using the politician's two-period objective from Equation 13, the payoff from delegating exceeds that from not delegating when offering benefit b^* if

$$\begin{aligned} &u_i(b_i^*, 1; e_t^*, \mu_t) + \pi_i u_i(b_i^*, 0; 0, \delta\mu_b^0(\pi_b)) \\ &> u_i(b_i^*, 0; 0, \mu_t) + \pi_i u_i(b_i^*, 0; 0, \delta\mu_t). \end{aligned} \quad (18)$$

For a first-term politician, delegation trades off between immediate and certain taxation costs and the possible benefit of future capacity enhancements upon reelection. She loses her constituents' share of the investment cost, or $\kappa_p e_t^*/2$. In the subsequent period, capacity increases to $\delta\mu_b^0(\pi_b)$ and reduces service times rather than depreciating to $\delta\mu_t$ and increasing service times.

Proposition 2 provides conditions under which delegation occurs. A group i politician will delegate when current capacity lies in an intermediate region denoted $\mathcal{D}_i$. Returning to the emergency response example from the introduction, costly improvements to police services will tend to occur when mayors, legislators, and police chiefs are relatively early in their terms and when service deficiencies are moderate. When response times are very high, investment is too expensive relative to the promise of future viability. And when response times are very low, either group i constituents would benefit little from further improvements or police departments are unwilling to invest further.

Proposition 2. Delegation. *Politician i delegates if and only if she is in her first term, the bureaucrat is of age* 1, *and*

$$\begin{aligned} \mu_t \in \mathcal{D}_i \equiv \Bigg(&\mu_b^0(\pi_b) + \frac{\pi_i}{\kappa_p}\left(2c\lambda_i\sqrt{\frac{\kappa_b}{\delta\pi_b m_b}} - \frac{\lambda_i^2 - k\Lambda^2}{\Lambda}\right), \\ &\min\left\{\frac{\Lambda}{\delta} + \frac{2\pi_i c\lambda_i}{\kappa_p}\sqrt{\frac{\kappa_b}{\delta\pi_b m_b}}, \mu_b^0(\pi_b)\right\}\Bigg). \end{aligned} \quad (19)$$

Interestingly, current viability is neither necessary nor sufficient for delegation. A politician may delegate to either viable or viable programs. She may even forego delegation to a viable program and render it unviable for the next period.

Figure 1 illustrates some basic relationships between capacity (μ_t), group demands (λ_1, λ_2), viability, and delegation for an insulated program. It holds total demand (Λ) constant, and thus lower group-1 demand corresponds to higher group-2 demand. There is a positive relationship between the size of the delegation region and constituent demands. Politicians representing a group with sufficiently high demand will always delegate to restore a low-capacity program, while others might give up. Highly unequal group demands can therefore produce steady, if uneven, political support over time.

The electoral context also affects delegation. An improvement in reelection prospects will typically increase the returns to delegation. Thus an increasing electoral advantage will expand one group's delegation region while shrinking the other's. In a similar fashion, an incumbency advantage will expand the delegation region for politicians of *both* groups. Proposition A.1 in the online appendix provides a result on the relationship between electoral prospects and delegation for insulated programs.

Finally, politicization does not fundamentally change delegation patterns. Compared with insulation, politicization reduces both the size of investments and their costs to politicians. This combination can cause a modest expansion of the delegation region. Figure A.1 in the online appendix provides an example comparing delegation regions under the two leadership systems.

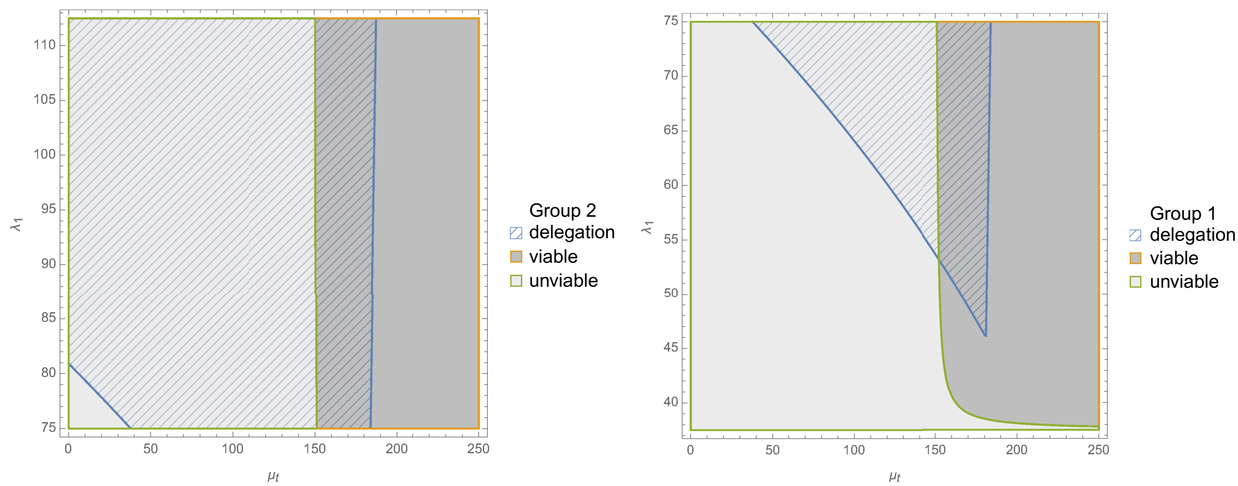
LONG-RUN SURVIVAL AND QUALITY

This section shows how institutional rules and political variables affect program survival and capacity. Governance quality over the long run depends on the joint evolution of the political environment and capacity. I use Markov chains to describe each of these processes.

The Markov chain denoted $\mathcal{P}_t$ summarizes the political environment. In each period, the political system is in a state represented by a triple (i, θ_i, θ_b), where $i \in \{1, 2\}$ is the group of the incumbent politician, and $\theta_i \in \{1, 2\}$ and $\theta_b \in \{1, 2\}$ are the term of the politician and the age of the bureaucrat, respectively. Under insulation, every

[9] Politicization could also plausibly imply bureaucrats with different policy preferences. Elected officials may select loyalists who reflect their preferences, or they may have restricted access to talent pools necessary for effective management. Since the incentive to invest depends on bureaucratic preferences, a politicized system may generate higher variation in bureaucratic investments. The qualitative results of such a model would remain similar to those developed here if the variation in bureaucratic preferences is moderate.

FIGURE 1. Inequality, Capacity, and Delegation for an Insulated Program



Note: Here, $\Lambda = \lambda_1 + \lambda_2 = 150$, $m_b = 75$, $c = 0.2$, $\kappa_b = 0.1$, $\kappa_p = 0.08$, $k = 0.0625$, $\pi_1 = 0.5$, and $\delta = 0.85$. Plots are of regions of delegation and program viability for a newly elected politician from group 1 (right) and 2 (left) as functions of capacity μ_t and demand rate λ_i. Because Λ is held constant, the vertical axes are linked, with high values of λ_2 in the left panel corresponding to lower values of λ_1 in the right panel. Note that group 2 is willing to delegate for arbitrarily low values of μ_t when $\lambda_2 > 81$.

combination of values is possible, and thus there are eight states. Under politicization, politician term and bureaucrat age coincide ($\theta_b = \theta_i$), and thus there are only four states. Between periods, the political system transitions to a new state with probability 0, π_1, or π_2, depending on the personnel system. Figure 2 represents the political processes under both structures, with nodes corresponding to political states.

The process $\mathcal{P}_t$ has straightforward properties. Under the assumed parameters of the game, every state is positive recurrent, and thus the process has a unique stationary distribution.[10] This implies that the long-run distribution of states is independent of the initial state. Conveniently, the probability of each state is easily calculated; Table 2 presents the stationary distribution for both leadership structures. Notably, since newly elected politicians always bring new bureaucrats, periods during which delegation may occur—that is, political states (1, 1, 1) and (2, 1, 1)—are twice as frequent under politicization. When neither party is electorally advantaged ($\pi_1 = \pi_2 = 1/2$), periods with age-1 bureaucrats and newly elected politicians occur with probability 2/3 under politicization and with probability 1/3 under insulation.

The frequency of delegation drives a central trade-off in determining program quality. By enabling more investment opportunities, politicization generates higher investment on the extensive margin. In contrast, Lemma 1 shows that insulation generates higher investment conditional upon delegation, or on the intensive margin. This trade-off fundamentally follows from the interaction of appointment rules and investment horizons, not by particular assumptions such as term length.

A second Markov chain, denoted $\mathcal{Q}_t$, describes capacity. Each state is represented by the four-tuple (i, θ_i, θ_b, μ), where i, θ_i, θ_b remain as before and μ is capacity. This Markov chain is infinite and produces much more complex paths than $\mathcal{P}_t$. As established, the incentives to delegate and invest depend on current capacity. At very low levels, neither group's politician may want to delegate and incur investment costs. Not delegating further erodes capacity and gives future politicians even less incentive to delegate. In this environment, capacity might converge to zero. By contrast, delegation and investment may persist indefinitely when starting capacity is sufficiently high.

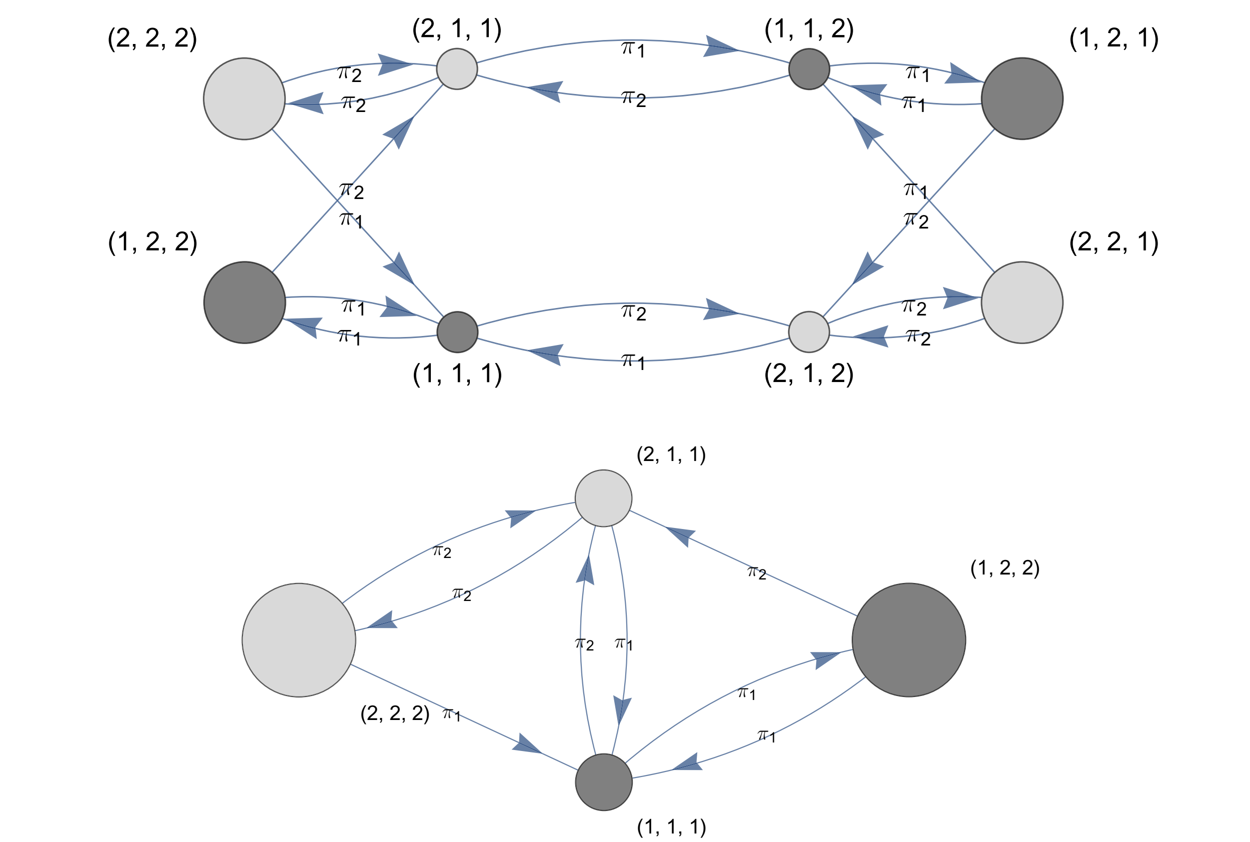
Program Survival

An important initial question is whether programs can survive over time. To illustrate, Figure 3 compares sample paths for capacity under different appointment structures. Two features stand out. First, under politicization, capacity occasionally depreciates but never falls very low. By comparison, under insulation capacity varies far more. Second, when inequality in group demands is low ($\lambda_1 = 74$, $\lambda_2 = 76$), an insulated program dies. But when inequality is higher ($\lambda_1 = 65$, $\lambda_2 = 85$), the program survives under both appointment structures, and in particular an insulated program can weather spells of low capacity.

The example in Figure 3 illustrates a more general point about the relationship between inequality and program persistence. As Figure 1 shows, when inequality is very low (i.e., λ_1 and λ_2 are close), neither group's politicians may be willing to rescue a low-capacity program. Under insulation, this situation becomes

[10] Under insulation, $\mathcal{P}_t$ has period 2 because of the fixed alternation of bureaucrats, and thus the distribution is stationary in the time average sense.

FIGURE 2. Bureaucratic Leadership Structures



Note: Vertices are political states, labeled (*politician group*, *politician term*, *bureaucrat age*); dark represents group 1, light represents group 2, small represents first-term politicians, and large represents second-term politicians. The top panel depicts an insulated program, where bureaucrats stay in office independently of the politician. The bottom panel depicts a politicized program, where bureaucrats leave office if their appointing politician loses reelection.

TABLE 2. Steady-State Distribution of Political States

State	Insulated	Politicized
(1, 1, 1)	$\frac{\pi_1}{2(1+\pi_1)}$	$\frac{\pi_1}{1+\pi_1}$
(1, 1, 2)	$\frac{\pi_1}{2(1+\pi_1)}$	…
(1, 2, 1)	$\frac{\pi_1^2}{2(1+\pi_1)}$	…
(1, 2, 2)	$\frac{\pi_1^2}{2(1+\pi_1)}$	$\frac{\pi_1^2}{1+\pi_1}$
(2, 1, 1)	$\frac{1-\pi_1}{2(2-\pi_1)}$	$\frac{1-\pi_1}{2-\pi_1}$
(2, 1, 2)	$\frac{1-\pi_1}{2(2-\pi_1)}$	…
(2, 2, 1)	$\frac{(1-\pi_1)^2}{2(2-\pi_1)}$	…
(2, 2, 2)	$\frac{(1-\pi_1)^2}{2(2-\pi_1)}$	$\frac{(1-\pi_1)^2}{2-\pi_1}$

possible after a series of periods with either an old bureaucrat or a second-term politician. Under politicization, matches between young politicians and young bureaucrats occur *at least every other period.* The resulting frequency of investment can prevent excessive decay.

Figure 1 also shows that when demand inequality is high, group 2 is willing to invest even if capacity is zero. This ensures that an insulated bureaucrat will eventually invest and rescue a program. An insulated program can therefore survive in a high inequality environment, despite occasional spells of very low capacity.

To be more precise about survival, I adopt the following two definitions. A transient program inevitably dies in the long run, as capacity is assured of falling to a level where no politician would delegate.

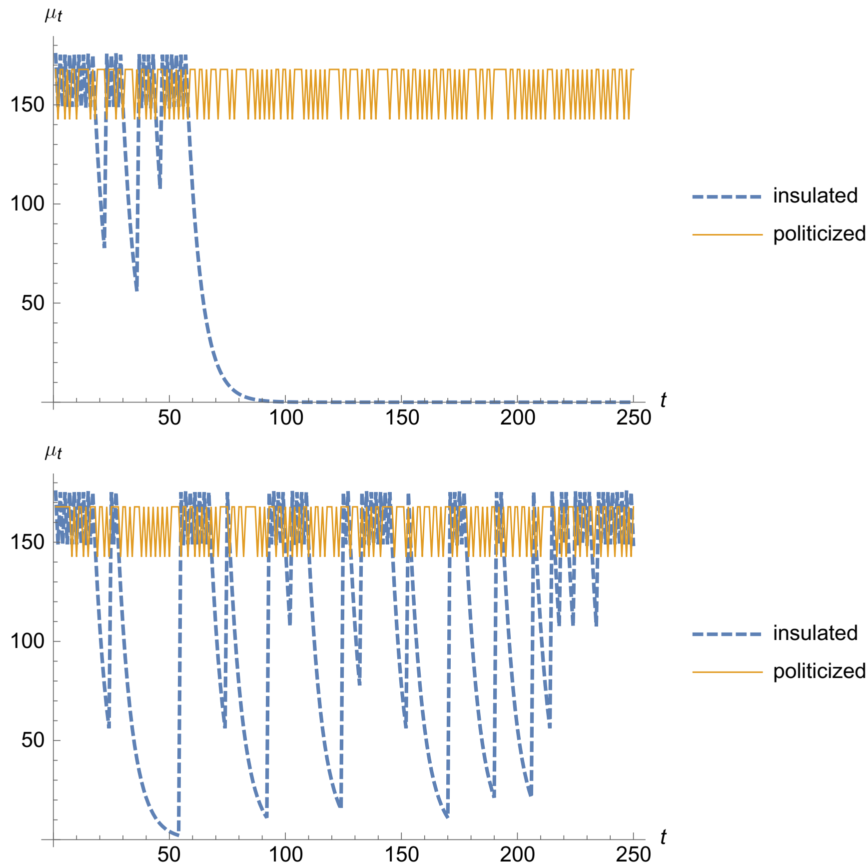
Definition 2. *A program is* transient *if* $\Pr\{\lim_{t\to\infty}\mu_t = 0\} = 1$.

Avoiding transience is possible if at least one politician is willing to delegate to programs of arbitrarily low capacity. In Figure 1, the group-2 politician fits this description when $\lambda_2 > 81$. The following definition formalizes this condition.

Definition 3. *A group i politician satisfies* deference *if*

$$\lambda_i > \underline{\lambda}_i(\pi_b) \equiv \Lambda\left(c\sqrt{\frac{\kappa_b}{\delta\pi_b m_b}} + \sqrt{\frac{c^2\kappa_b}{\delta\pi_b m_b} + \frac{\kappa_p}{\pi_i}\left(\frac{1}{\delta} + \frac{1}{\Lambda}\sqrt{\frac{\pi_b m_b}{\delta\kappa_b}}\right) + k}\right). \tag{20}$$

FIGURE 3. Capacity Paths and Inequality



Note: Here, $\Lambda = \lambda_1 + \lambda_2 = 150$, $m_b = 75$, $c = 0.2$, $\kappa_b = 0.1$, $\kappa_p = 0.08$, $k = 0.0625$, $\pi_1 = 0.5$, and $\delta = 0.85$. Both panels show μ_t over 250 periods under insulation and politicization. At top, $\lambda_1 = 74$; at bottom, $\lambda_1 = 65$.

Deference does not always require a politician to delegate; by Proposition 2, she will not delegate when capacity is very high because no investment would result. A politician who satisfies deference is in effect willing to start an entirely new program where none exists ($\mu_t = 0$). Without such a politician, avenues for program creation might include shocks to a group's demand or the repurposing of an existing program or spare capacity.

The next result provides conditions under which programs survive indefinitely. Both insulated and politicized programs avoid transience if at least one group's politicians satisfy deference. Politicized programs have the additional advantage of surviving when *both* groups are only moderately willing to invest. In the emergency response setting, the former condition corresponds to a community where one group deems police responses as essential and worth rescuing even at great cost. The latter condition reflects a more homogeneous community where group demands are more modest, due perhaps to lower crime rates.

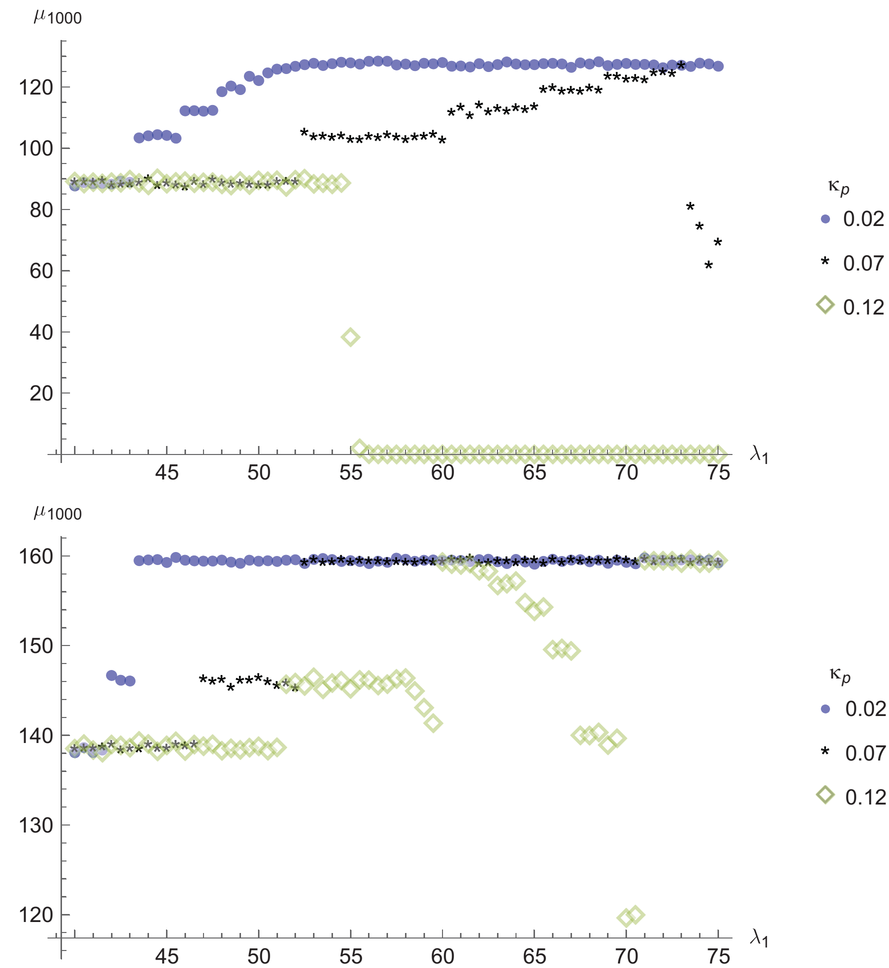
Proposition 3. Program Survival. *(i) An insulated program is not transient if and only if some group satisfies deference.*

(ii) A politicized program is not transient if either some group satisfies deference, or $\mu_1 \in \mathcal{D}_i$ for the period-1 incumbent of group i and $\lambda_i \geq \underline{\lambda}_i^p$ for both groups, where $\underline{\lambda}_i^p$ is the minimum value of λ_i such that $\delta^2\mu_b^0(\pi_1) \in \mathcal{D}_i$ and $\delta^2\mu_b^0(\pi_2) \in \mathcal{D}_i$.

The intuition for part (i) of Proposition 3 is that if neither group is willing to rescue a sufficiently low-capacity program, then under insulation a path of electoral outcomes that ensures complete deterioration will eventually occur with certainty. Such a path consists of a sufficiently long series of either newly elected politicians coupled with age-2 bureaucrats or reelected politicians coupled with age-1 bureaucrats. Players along this path allow so much depreciation that eventually no newly elected politician would ever reinitiate investment, even when matched with a young bureaucrat. As an example, this path becomes possible in Figure 1 when λ_1 and λ_2 are close to 75.

Part (ii) shows that politicization is particularly conducive to survival when inequality is low. In addition to the logic of part (i), politicization adds two factors that facilitate survival. First, it can eliminate lengthy episodes without investment. Due to the frequency of pairings between first-term politicians and age-1

FIGURE 4. Long-Run Capacity, Insulated and Politicized Programs



Note: Here, $\Lambda = \lambda_1 + \lambda_2 = 150$, $c = 0.2$, $\kappa_b = 0.1$, $k = 0.0625$, $\pi_1 = 0.5$, $\mu_1 = 105$, $m_b = 75$, and $\delta = 0.85$. Both panels depict average μ_{1000} across 5,000 simulations as a function of λ_1, varying κ_p. The top panel plots an insulated program, and the bottom panel plots a politicized program.

bureaucrats, newly elected politicians are continuously able to mold programs to fit their needs. Persistence is assured if initial capacity is high and the delegation region is large enough to withstand just two rounds of noninvestment and depreciation.[11] Second, the lower investments of politicized bureaucrats (expressed by the threshold $\underline{\lambda}_i(\pi_b)$ in Equation 20) reduce the cost of delegation. Importantly, these factors alone do not guarantee survival: once a group becomes unwilling to delegate, its repeated reelection will eventually lead to complete depreciation.

Figure 4 numerically illustrates the effect of inequality on survival and capacity for insulated and politicized programs. Each point is the average terminal capacity level at period 1,000 (μ_{1000}) over 5,000 simulation runs at different values of group-1 demand (λ_1). As before, holding total demand (Λ) constant, inequality increases when λ_1 decreases and λ_2 increases.

The top panel of Figure 4 clearly shows that insulated programs can become transient as inequality decreases (i.e., λ_1 and λ_2 are close). When politicians face high investment costs (κ_p), capacity drops all the way to zero and stays there. By contrast, politicized programs always survive in this example. A more unequal society prevents transience by guaranteeing that at least one group will be willing to delegate to a low-capacity program.

Proposition 3 finally has implications for how the electoral environment affects survival. Corollary 1 shows that as a group's electoral prospects improve, the deference condition becomes easier to satisfy and it becomes more willing to sustain programs. Figure A.3 in the online appendix shows that when investment costs are relatively high, capacity for an insulated program drops to zero when group 2 is electorally disadvantaged. Since this group has higher demand, an electoral disadvantage weakens the more important source of investment. Thus in addition to inequality, an electoral tilt in favor of high-demand groups can ensure survival. The result follows from straightforward differentiation of the threshold for

[11] The proof of the result provides closed-form expressions for the values of λ_i required for a delegation region of this size.

deference $\underline{\lambda}_i(\pi_b)$ (Equation 20), and is stated without proof.

Corollary 1. *Electoral Prospects and Transience.* $\underline{\lambda}_i(\pi_b)$ *is decreasing in* π_i.

Somewhat surprisingly, politicized programs may also become more survivable as group 1's (the low-demand group) electoral prospects improve. This happens because of the mechanism in Proposition 3(ii), whereby both groups are willing to maintain a politicized program. Electoral conditions are therefore not clearly critical for the survival of such programs.

Long-Run Quality

Beyond survival, citizens should also care about the quality of programs as they operate. Figure 4 illustrates several features of program quality. For example, higher politician costs (κ_p) reduce the delegation region and hence average capacity.[12] Higher costs can also reduce capacity by preventing programs from achieving long-run sustainability. This can occur if the low-demand group wins the first few elections and allows enough depreciation to induce early program failure.[13] This possibility produces nonmonotonicities in average capacity as inequality declines. Along with the conditions for transience from the previous subsection, early failures are another reason why inequality does not necessarily harm program quality.

Perhaps most interestingly, the politicized program in the figure performs better than the insulated program. Empirical studies in different contexts have produced some support for the benefits of political control (e.g., Raffler 2019), but also significant evidence in favor of insulating reforms such as civil service policies (e.g., Carpenter 2001; Lewis 2007; Rauch 1995). The trade-off in the model between the rate of investment under politicization and the intensity of investment under insulation provides a basis for comparing the relative benefits of these structures.

A prominent example of police reform illustrates how political control over personnel can drive service improvements. Amidst a surge in national attention toward police violence in 2020, the city of Camden, New Jersey, stood out for its improving police–community relations and decreasing levels of police violence and serious crime. The department achieved some of its success through policy initiatives that changed use of force guidelines and reduced wait times.[14] A unique combination of personnel moves enabled these reforms. Between 2002 to 2010, New Jersey administered the city under emergency powers, which it used in 2008 to appoint Scott Thomson as one of the state's youngest police chiefs. At the behest of the governor and mayor, Thomson led a 2012 dismantling and reconstitution of the department, which proceeded under a temporary suspension of some civil service hiring rules. The police force thus functioned without much of the political insulation found in its counterparts and instead followed its political leadership by instituting consequential reforms.[15]

The next results compare the long-run program quality of nontransient programs. (Because their capacity depreciates to zero with certainty, any transient program has a long-run average capacity of zero.) Proposition 4 reports average capacity, for which analytical solutions are fortunately possible under some mild assumptions. The result immediately implies that neither leadership structure performs unambiguously better.

Proposition 4. Long-Run Quality. *Suppose that both groups satisfy deference, Λ is sufficiently high, and $\mu_1 \in \mathcal{D}_i$ for the period 1 incumbent of group i.*

(i) *An insulated program's average capacity is*

$$\sum_{i=1}^{2} \frac{(1+\delta)\left(1/2+\pi_1\pi_2-\delta^2\pi_i^3(1+\pi_{-i})\right)\left(\Lambda+\sqrt{\delta m_b/\kappa_b}\right)}{2(1+\pi_1)(1+\pi_2)\left(1-\delta^2(1-2\pi_1\pi_2)\right)}. \tag{21}$$

(ii) *A politicized program's average capacity is*

$$\sum_{i=1}^{2} \frac{\pi_i(1+\delta\pi_i)\left(\Lambda+\sqrt{\delta\pi_i m_b/\kappa_b}\right)}{1+\pi_i}. \tag{22}$$

In addition to deference, the result requires the potential client population (Λ) to be sufficiently high. This ensures that politicians of both groups delegate whenever bureaucrats are willing to make positive investments (i.e., for any any capacity level $\mu_t < \mu_b^0(\pi_b)$). In other words, politicians do not allow depreciation from the threshold $\mu_b^0(\pi_b)$ in Equation 15. This produces a simple evolutionary trajectory whereby capacity depreciates in every period until the political system reaches state (1, 1, 1) or (2, 1, 1), where newly elected politicians delegate to young bureaucrats. The condition also holds if the politician's marginal costs (κ_p) are low or depreciation (δ) is slow.[16]

[12] Figure A.2 in the online appendix shows a similar relationship for the bureaucratic cost parameter κ_b.

[13] One interesting feature of Figure 4 is the presence of kinks in average capacity as λ_1 increases. This is due in part to changes in the size of the delegation region that allow more periods of noninvestment before a politician who is willing to delegate is elected.

[14] See Jen A. Miller, "How Tech Can Help Cities Reduce Crime," *CIO*, April 9, 2014, and "NJ Should Be Proud of Camden Police Reform," *New Jersey Law Journal*, July 27, 2020.

[15] See James Osborne, "N.J. Civil Service Panel's Ruling Boosts New Camden Police Force," *Philadelphia Inquirer*, October 4, 2012, and Anne Milgram, "The Camden Policing Model," *cafe.com* audio post, June 18, 2020, retrieved December 1, 2020. According to Milgram, the New Jersey attorney general who first appointed Thomson, "the civil service rules have stopped … innovation from flourishing in departments. The most innovative officers are not the ones who are promoted, and usually people are promoted at the end of their career to be chief." As of November 2020, about half of police chiefs in New Jersey cities with population exceeding 50,000 had served for fewer than three years.

[16] The online appendix defines the applicable *full deference* condition. The condition holds if $\Lambda > \sqrt{\frac{\delta^3 m_b}{\kappa_b(1-\delta)^2}}$.

Corollary 2 uses Proposition 4 to derive potentially testable relationships between the electoral environment, program management, and long-run quality. Part (i) establishes that insulated programs benefit from an unbiased electorate (i.e., $\pi_1 = 1/2$). This happens because a competitive electorate maximizes the chances of the political states that generate investment. By contrast, part (ii) shows that politicized programs do not necessarily benefit from an unbiased electorate. A competitive electorate continues to help when durability is low (i.e., low δ), but such programs may perform better in skewed electorates when durability is high.

Part (iii) shows that the superiority of politicization illustrated in Figure 4 is not fully general: insulated programs gain an advantage when the electorate is unbiased and programs are durable. Recall that the weakness of insulation was infrequent investment that led to depreciation. As in part (i), a competitive electorate maximizes the frequency of delegation. In addition, high durability mitigates the effects of periods without delegation. In conjunction these factors allow the higher investment levels under insulation to generate better performance.

Corollary 2. Politicized versus Insulated. *Suppose that both groups satisfy the conditions of Proposition 4 under both insulated and politicized programs.*

(i) *Under insulation, average program quality is maximized at* $\pi_1 = 1/2$.
(ii) *Under politicization, there exists* $\delta_p \in (0, 1)$ *such that average program quality is maximized at* $\pi_1 = 1/2$ *only if* $\delta \leq \delta_p$.
(iii) *For* δ *sufficiently near* 1, *average program quality is higher under insulation. For* $\pi_1 = 1/2$, *there exists* $\widehat{\delta} \in (0, 1)$ *such that average program quality is higher under insulation if and only if* $\delta > \widehat{\delta}$.

Empirical results in different contexts have generated different answers about the direction of the relationship between electoral security and program performance (e.g., Pepinsky, Pierskalla, and Sacks 2017; Vakilifathi 2019). These results usefully provide distinct conditions under which both competitive and uncompetitive electorates are beneficial. As further examples, Figure A.3 in the online appendix shows how, consistent with Corollary 2, an unbiased electorate maximizes capacity when an insulated program is assured of long-run survival. However, as the previous subsection discusses, an electoral advantage in favor of the high-demand group is sometimes necessary to assure survival in the first place.

Returning to an example from the introduction, a variety of appointment procedures govern the leadership of U.S. municipal police forces. Some are directly elected, while others are appointed by mayors or nonpolitical city managers. Corollary 2 predicts that the most insulated chiefs would perform best on metrics such as reducing wait times in an electorally competitive environment. This would not necessarily be the case for department heads who are more exposed to electoral pressures. American infrastructure projects, which are administered by local or regional officials with highly heterogeneous appointment structures, offer an arena for examining service quality when capacity is durable (Gerber and Gibson 2009). Part (iii) of the corollary predicts that insulated programs would outperform politicized ones in electorally competitive regions.

It is finally worth observing that plausible alternative assumptions could tilt the trade-offs in appointment structures toward insulation. Longer career horizons would raise the investment incentives of insulated bureaucrats more than those of politicized bureaucrats (who expect short careers in a competitive electorate). The ability to gain proficiency with experience would also improve the performance of programs with longer-lived managers. Nevertheless, even without these extensions, the model provides an account of the relative advantages of each type of leadership structure.

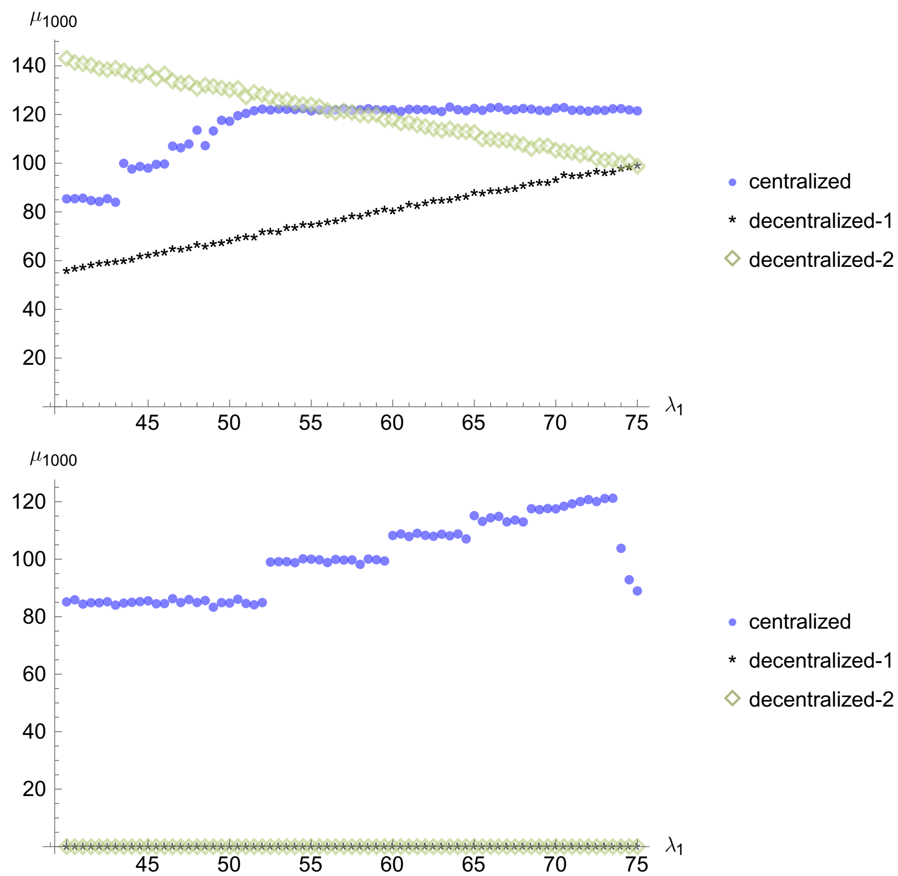
Extension: Decentralization

Queues can serve as a basis for considering service provision in a range of institutional settings. One important example is the centralization and devolution of local services, such as the U.S. movement toward consolidating municipal police departments over the past few decades (Wilson et al. 2018). The performance implications of centralization are not obvious. A classic trade-off from the study of federalism is that decentralization may increase local experimentation and learning but exacerbate externalities (e.g., Bednar 2011). Extending the model to incorporate decentralized servicers allows it to show how capacity can play an important role in determining the allocation of authority.

I modify the model so that each group runs and pays for its own independent service. All politicians within a group are identical and maximize their group's welfare. Finally, the group-i bureaucrat's public-service motivation is set equal to its demand rate ($m_b = \lambda_i$). This combination of features enables a close comparison with the basic model.

Figure 5 numerically compares centralized and decentralized capacity for an insulated program at different values of politician marginal cost (κ_p), using the same parameters as Figure 4 where possible. The main comparison of interest is between the *centralized* regime of the basic model and the decentralized units labeled *decentralized-1* and *decentralized-2*. Two effects of centralization are immediately clear. First, when investment costs are low, both decentralized units are independently able to sustain programs indefinitely and the resulting service is largely superior to that under centralization. Despite serving only half of the population, local capacity levels sometimes even exceed those of a centralized provider. Second, when investment costs are high, the smaller units are unable to sustain programs even while the center can. This occurs because decentralized politicians cannot pass some investment costs onto opposition voters. Thus, a centralized authority has advantages in capacity

FIGURE 5. Long-Run Capacity and Decentralization (Insulated Program)



Note: Here, $\Lambda = \lambda_1 + \lambda_2 = 150$, $m_b = 75$, $c = 0.2$, $\kappa_b = 0.1$, $k = 0.0625$, $\pi_1 = 0.5$, $\mu_1 = 105$, and $\delta = 0.85$. Both panels depict average μ_{1000} across 5,000 simulations as a function of λ_1 for centralized and decentralized systems. In the top panel, $\kappa_p = 0.02$; in the bottom, $\kappa_p = 0.07$.

building that may outweigh the benefits of decentralizing control and tailoring services to local conditions.

CONCLUSIONS

In recent years, political scientists have increasingly regarded service provision as a critical output, both for its social welfare implications and its ability to illuminate political processes. For many reasons, queues are a natural conceptual starting point for analyzing service settings. Queues are pervasive in both physical and virtual forms, and especially so in the public sector. They have become common formal and informal metrics for measuring organizational performance. Finally, a large body of queueing models already exists.

This paper connects queueing theory and governance quality. In the model, queues link citizens with the bureaucracy in a simple but dynamic political economy framework. In this framework, political actors are fully strategic: politicians choose both policy and delegation in the shadow of reelection, while public-service-motivated bureaucrats choose investments under different appointment rules. The combination of these elements allows the theory to incorporate many now-standard features of political economy models of the bureaucracy.

The model identifies several challenges in sustaining agency capacity. In addition to group demands, capacity-enhancing investments require the confluence of willing politicians and bureaucrats. This produces two unexpected benefits of partisan inequality and politicization. Inequality produces constituencies who are always willing to support—or rescue—programs. And despite the reduced incentives to invest brought on by electoral uncertainty, politicization provides regular political willpower that can increase long-run quality. The results have the potential to address a range of empirical questions about the determinants of service quality, including the roles of leadership structures, the electoral environment, and decentralization.

The framework provides many openings for further inquiry. On the politician side, there are clear incentives to design policies that either discriminate in costs or benefits or manipulate program eligibility. Perversely, programs might become more survivable if low-demand politicians find ways to reduce enrollment through eligibility restrictions or discrimination. On the citizen side, voting could discipline politicians, while also possibly introducing other distortions. Behaviors such as bribery or exiting queues and more realistic queues that allow multiple servicers, observable lengths, pricing, or privatization would add to the richness of service provision. Even more ambitiously, a more sophisticated service provider could adjust her solution process in the presence of impatient or politically influential constituents. The broader implication is that queues provide a tractable foundation for analyzing the relationships between political systems and the citizens they serve.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0003055421000046.

ACKNOWLEDGMENTS

I thank Daniel Carpenter, Jean Guillaume Forand, Michael Herron, Alex Hirsch, Ryan Hubert, Giovanna Invernizzi, Jacob Lew, David Simpson, Tara Slough, Ian Turner, Gergely Ujhelyi, and Anna Wilke for helpful comments. Seminar audiences at the University of Gothenburg, Bocconi University, NYU, the University of Wisconsin, and panel participants at the 2017 APSA conference, 2018 Emory Conference on Institutions and Law-Making, 2018 WPSA conference, 2018 Yale CSAP conference, and 2019 American Law and Economics Association conference also provided useful feedback.

REFERENCES

Aberbach, Joel D., and Bert A. Rockman. 2000. *In the Web of Politics: Three Decades of the U.S. Federal Executive*. Washington, DC: Brookings Institution.

Akhtari, Mitra, Diana Moreira, and Laura Trucco. 2020. "Political Turnover, Bureaucratic Turnover, and the Quality of Public Services." Unpublished Manuscript. University of California, Davis.

Ando, Amy W. 1999. "Waiting to be Protected under the Endangered Species Act: The Political Economy of Regulatory Delay." *Journal of Law and Economics* 42 (1): 29–60.

Asher, Jeff. 2018. "Fewer Crimes Get Counted When Police Are Slow To Respond." *Fivethirtyeight.com.* Retrieved Feb. 23, 2020. https://fivethirtyeight.com/features/fewer-crimes-get-counted-when-police-are-slow-to-respond.

Banerjee, Abhijit V. 1997. "A Theory of Misgovernance." *Quarterly Journal of Economics* 112 (4): 1289–332.

Bednar, Jenna. 2011. "The Political Science of Federalism." *Annual Review of Law and Social Science* 7 (1): 269–88.

Beggs, A. W. 2001. "Queues and Hierarchies." *Review of Economic Studies* 68 (2): 297–322.

Berry, Christopher R., Barry C. Burden, and William G. Howell. 2010. "After Enactment: The Lives and Deaths of Federal Programs." *American Journal of Political Science* 54 (1): 1–17.

Besley, Timothy, and Torsten Persson. 2009. "The Origins of State Capacity: Property Rights, Taxation, and Politics." *American Economic Review* 99 (4): 1218–44.

Bilmes, Linda J., and Jeffrey R. Neal. 2003. "The People Factor: Human Resources Reform in Government." In *For the People: Can We Fix Public Service?* eds. John D. Donahue and Joseph S. Nye Jr., 113–33. Washington, DC: Brookings Institution.

Blanes i Vidal, Jordi, and Tom Kirchmaier. 2018. "The Effect of Police Response Time on Crime Clearance Rates." *Review of Economic Studies* 85 (2): 855–91.

Bolton, Alexander, Rachel Augustine Potter, and Sharece Thrower. 2016. "Organizational Capacity, Regulatory Review, and the Limits of Political Control." *Journal of Law, Economics, & Organization* 32 (2): 242–71.

Bussell, Jennifer. 2019. *Clients and Constituents: Political Responsiveness in Patronage Democracies.* New York: Oxford University Press.

Callander, Steven, and Gregory J. Martin. 2017. "Dynamic Policymaking with Decay." *American Journal of Political Science* 61 (1): 50–67.

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

Carpenter, Daniel P. 2002. "Groups, the Media, Agency Waiting Costs, and FDA Drug Approval." *American Journal of Political Science* 46 (3): 490–505.

Carpenter, Daniel P. 2004. "Commentary: Staff Resources Speed FDA Drug Review: A Critical Analysis of the Returns to Resources in Approval Regulation." *Journal of Health Politics, Policy and Law* 29 (3): 431–42.

Carpenter, Daniel P., and David E. Lewis. 2004. "Political Learning from Rare Events: Poisson Inference, Fiscal Constraints, and the Lifetime of Bureaus." *Political Analysis* 12 (3): 201–32.

Dahlström, Carl, Victor Lapuente, and Jan Teorell. 2012. "The Merit of Meritocratization: Politics, Bureaucracy, and the Institutional Deterrents of Corruption." *Political Research Quarterly* 65 (3): 656–68.

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): 1169–218.

Derthick, Martha. 1990. *Agency Under Stress: The Social Security Administration in American Government.* Washington, DC: The Brookings Institution.

Downs, Anthony. 1967. *Inside Bureaucracy.* Boston: Little Brown.

Epstein, David, and Sharyn O'Halloran. 1999. *Delegating Powers: A Transaction Cost Politics Approach to Policy Making Under Separate Powers.* New York: Cambridge University Press.

Francois, Patrick. 2000. "Public Service Motivation as an Argument for Government Provision." *Journal of Public Economics* 78 (3): 275–99.

Gailmard, Sean, and John W. Patty. 2007. "Slackers and Zealots: Civil Service, Policy Discretion and Bureaucratic Expertise." *American Journal of Political Science* 51 (4): 873–89.

Gailmard, Sean, and John W. Patty. 2012. "Formal Models of Bureaucratic Politics." *Annual Review of Political Science* 15: 353–77.

Gerber, Elisabeth R., and Clark C. Gibson. 2009. "Balancing Regionalism and Localism: How Institutions and Incentives Shape American Transportation Policy." *American Journal of Political Science* 53 (3): 633–48.

Gratton, Gabriele, Luigi Guiso, Claudio Michelacci, and Massimo Morelli. 2020. "From Weber to Kafka: Political Instability and the Overproduction of Laws." Unpublished Manuscript. Bocconi University.

Gross, Donald, John F. Shortle, James M. Thompson, and Carl M. Harris. 2008. *Fundamentals of Queueing Theory* (4th edition). Hoboken, NJ: John Wiley and Sons.

Heclo, Hugh. 1977. *A Government of Strangers: Executive Politics in Washington.* Washington, DC: The Brookings Institution.

Herd, Pamela, and Donald P. Moynihan. 2018 *Administrative Burden: Policymaking by Other Means.* New York: Russell Sage.

Herron, Michael C., and Daniel A. Smith. 2016. "Precinct Resources and Voter Wait Times." *Electoral Studies* 42: 249–63.

Hirsch, Alexander V., and Kenneth W. Shotts. 2012. "Policy-Specific Information and Informal Agenda Power." *American Journal of Political Science* 56 (1): 67–83.

Hollibaugh, Gary E., Gabriel Horton, and David E. Lewis. 2014. "Presidents and Patronage." *Journal of Politics* 58 (4): 1024–42.

Horn, Murray J. 1995. *The Political Economy of Public Administration.* New York: Cambridge University Press.
Huber, John D., and Nolan M. McCarty. 2004. "Bureaucratic Capacity, Delegation, and Political Reform." *American Political Science Review* 98 (3): 481–94.
Inter-American Development Bank. 2018. *Wait No More: Citizens, Red Tape, and Digital Government.* eds. Benjamin Roseth, Angela Reyes, Carlos Santiso. Washington, DC: Inter-American Development Bank.
Kaufman, Herbert. 1976. *Are Government Organizations Immortal?* Washington, DC: Brookings Institution Press.
Keefer, Philip, and Stuti Khemani. 2005. *Democracy, Public Expenditures, and the Poor: Understanding Political Incentives for Providing Public Services.* Washington, DC: World Bank Group.
Krause, George A., David E. Lewis, and James W. Douglas. 2006. "Political Appointments, Civil Service Systems, and Bureaucratic Competence: Organizational Balancing and Executive Branch Revenue Forecasts in the American States." *American Journal of Political Science* 50 (3): 770–87.
Kruks-Wisner, Gabrielle. 2018. *Claiming the State: Active Citizenship and Social Welfare.* New York: Cambridge University Press.
Lewis, David E. 2007. "Testing Pendleton's Premise: Do Political Appointees Make Worse Bureaucrats?" *Journal of Politics* 69 (4): 1073–88.
Lipsky, Michael. 1980. *Street-Level Bureaucracy: The Dilemmas of the Individual in Public Service.* New York: Russell Sage Foundation.
Lui, Francis T. 1985. "An Equilibrium Queueing Model of Bribery." *Journal of Political Economy* 93 (4): 760–81.
Maranto, Robert. 1998. "Thinking the Unthinkable in Public Administration: A Case for Spoils in the Federal Bureaucracy." *Administration and Society* 29 (6): 623–42.
Moe, Terry M. 1982. "Regulatory Performance and Presidential Administration." *American Journal of Political Science* 26 (2): 197–224.
Moe, Terry M. 1985. "The Politicized Presidency." In *The New Direction in American Politics*, eds. John E. Chubb and Paul E. Peterson, 235–71. Washington, DC: Brookings Institution.
Moe, Terry M. 1989. "The Politics of Bureaucratic Structure." In *Can the Government Govern?* eds. John E. Chubb and Paul E. Peterson, 267–329. Washington, DC: Brookings Institution.
Mueller, Hannes. 2015. "Insulation or Patronage: Political Institutions and Bureaucratic Efficiency." *The B.E. Journal of Economic Analysis & Policy* 15 (3): 961–96.
Naor, P. 1969. "The Regulation of Queue Size by Levying Tolls." *Econometrica* 37 (1) 15–24.
Nath, Anusha. 2015. "Bureaucrats and Politicians: How Does Electoral Competition Affect Bureaucratic Performance?" IED Working Paper 269.
Pepinsky, Thomas B., Jan H. Pierskalla, and Audrey Sacks. 2017. "Bureaucracy and Service Delivery." *Annual Review of Political Science* 20: 249–68.
Prendergast, Canice. 2003. "The Limits of Bureaucratic Efficiency." *Journal of Political Economy* 111 (5): 929–58.
Raffer, Pia. 2019. "Does Political Oversight of the Bureaucracy Increase Accountability? Field Experimental Evidence from an Electoral Autocracy." Unpublished Manuscript. Harvard University.
Rainey, Hal G., and Paula Steinbauer. 1999. "Galloping Elephants: Developing Elements of a Theory of Effective Government Organizations." *Journal of Public Administration Research and Theory* 9 (1): 1–32.
Rauch, James E. 1995. "Bureaucracy, Infrastructure, and Economic Growth: Evidence from U.S. Cities During the Progressive Era." *American Economic Review* 85 (4): 968–79.
Rogger, Daniel. 2018. "The Consequences of Political Interference in Bureaucratic Decision Making: Evidence from Nigeria." World Bank Development Economics Research Group, Policy Research Working Paper 8554.
Romig, Kathleen. 2017. "More Cuts to Social Security Administration Funding Would Further Degrade Service." Center on Budget and Policy Priorities. Retrieved Feb. 23, 2020. https://www.cbpp.org/research/social-security/cuts-weakening-social-security-administration-services.
Rosen, Stephen P. 1988. "New Ways of War: Understanding Military Innovation." *International Security* 13 (1): 134–68.
Slough, Tara. 2020. "Bureaucrats Driving Inequality in Access: Experimental Evidence from Colombia." Unpublished Manuscript. New York University.
Ting, Michael M. 2017. "Politics and Administration." *American Journal of Political Science* 61 (2): 305–19.
Ting, Michael M., Olle Folke, Shigeo Hirano, and James Snyder. 2013. "Elections and Reform: The Adoption of Civil Service Systems in the U.S. States." *Journal of Theoretical Politics* 25 (3): 363–87.
Turner, Ian. 2019. "Political Agency, Oversight, and Bias: The Instrumental Value of Politicized Policymaking." *Journal of Law, Economics, & Organization* 35 (3): 544–78.
Ujhelyi, Gergely. 2014. "Civil Service Reform." *Journal of Public Economics* 118: 15–25.
United States Government Accountability Office. 2014. "Elections: Observations on Wait Times for Voters on Election Day 2012." Report GAO-14-850.
United States Government Accountability Office. 2015. "VA Mental Health: Action Needed to Improve Access Policies and Wait-Time Data." Report GAO-16-170T.
United States Government Accountability Office. 2018. "Aviation Security: TSA Uses Data to Monitor Airport Operations and Respond to Increases in Passenger Wait Times and Throughput." Report GAO-18-563T.
United States Social Security Administration. 2018. "Fiscal Year 2018 Budget Overview." Retrieved November 30, 2020. https://www.ssa.gov/budget/FY18Files/2018BO.pdf.
Vakilifathi, Mona. 2019. "Do Electorally-Vulnerable Legislators Grant More or Less Statutory Discretion?" Unpublished Manuscript. New York University.
Whitford, Andrew B. 2005. "The Pursuit of Political Control by Multiple Principals." *Journal of Politics* 67 (1): 28–49.
Wilke, Anna. 2020. "How Does the State Replace the Community? Experimental Evidence on Crime Control from South Africa." Unpublished Manuscript. Columbia University.
Wilson, Jeremy M., Steve Chermak, Nicholas Corsaro, Clifford Grammich, and Jeffrey Gruenewald. 2018. "A Multi-Site Assessment of Police Consolidation: Final Summary Overview." U.S. National Criminal Justice Reference Service Report 252503.

APPENDIX

Proof of Proposition 1. Since delegation and investment can only affect future capacity, there is no delegation if either the politician or bureaucrat is in her terminal period of office.

Using Equation 7 and Equation 9, citizens will join the queue with probability 1 if

$$\frac{\lambda_i}{\Lambda} \geq \frac{c}{\mu - \Lambda}. \tag{23}$$

Next, using Equation 3, the politician will prefer providing benefits ($s = 1$) using this solution to closing down the program ($s = 0$) if the following condition holds:

$$\lambda_i\left(b_i^*-\frac{c}{\mu-\Lambda}\right)-\frac{\Lambda\left(k+b_i^{*2}\right)}{2}>0 \tag{24}$$

$$\frac{\lambda_i}{\Lambda}>\frac{\Lambda k}{\lambda_i}+\frac{2c}{\mu-\Lambda}. \tag{25}$$

Equation 25 implies Equation 23, and is thus sufficient for ensuring a pure strategy queueing equilibrium. Solving for μ produces the expression for $\underline{\mu}_i$ in Equation 10. Thus, for $\mu \leq \underline{\mu}_i$ the politician can receive no more than 0 and chooses s^* = 0. Otherwise she chooses s^* = 1 and b^* as derived in Equation 9.

The politician's expected utility from a single period without delegation can be found by substituting these values into the politician's objective given in Equation 3. ■

Proof of Lemma 1. First consider the politician's choice of s_t and b_t. Note that the only effect of any investment e_t on the politician's maximization problem is through period t taxes that are independent of s_t and b_t. Since s_t and b_t also do not affect period t + 1 payoffs, her optimization problem from Equation 13 is identical to her one-period maximization problem. Thus the politician's optimal policies are given by s_i^* and b_i^* in Equation 9, as derived in Proposition 1.

For an age-1 bureaucrat's investment decision, the first-order condition of Equation 14 is

$$\frac{\delta\pi_b m_b}{(\delta(e_t+\mu_t)-\Lambda)^2}-\kappa_b=0.$$

The second derivative is

$$-\frac{2\delta^2\pi_b m_b}{(\delta(e_t+\mu_t)-\Lambda)^3}.$$

Since $\delta(e_t+\mu_t)>\Lambda$ at any utility-maximizing solution, this is clearly negative.

Solving the first-order condition for e then produces the optimal interior investment level.

$$e^*=\mu_b^0(\pi_b)-\mu_t.$$

At a corner solution, this value is negative and e^* = 0. ■

Proof of Proposition 2. I begin by calculating the politician's net benefit of delegation for different values of μ_t. There are three cases. First, when $\mu_t>\mu_b^0(\pi_b)$, the bureaucrat's optimal investment is 0, and there is no benefit from delegation.

Second, when optimal investment is positive and the program would remain viable without investment ($\mu_t>\underline{\mu}_i/\delta$), substituting into Equation 18 produces the interior net benefit of delegation:

$$\bar{\vartheta}(\mu_t)=\pi_i\lambda_i c\left(\frac{1}{\delta\mu_t-\Lambda}-\sqrt{\frac{\kappa_b}{\delta\pi_b m_b}}\right)+\frac{\kappa_p}{2}\left(\mu_t-\frac{\Lambda}{\delta}-\sqrt{\frac{\pi_b m_b}{\delta\kappa_b}}\right). \tag{26}$$

This function has roots at $\mu_b^0(\pi_b)$ and $\frac{\Lambda}{\delta}+\frac{2\pi_i c\lambda_i}{\kappa_p}\sqrt{\frac{\kappa_b}{\delta\pi_b m_b}}$. Furthermore, it is strictly convex for $\mu_t>\Lambda/\delta$ and positive only if $\mu_t>\Lambda/\delta$. Define the following values:

$$\hat{\mu}_i^-=\min\left\{\mu_b^0(\pi_b),\frac{\Lambda}{\delta}+\frac{2\pi_i c\lambda_i}{\kappa_p}\sqrt{\frac{\kappa_b}{\delta\pi_b m_b}}\right\},\ \text{and} \tag{27}$$

$$\hat{\mu}_i^+=\max\left\{\mu_b^0(\pi_b),\frac{\Lambda}{\delta}+\frac{2\pi_i c\lambda_i}{\kappa_p}\sqrt{\frac{\kappa_b}{\delta\pi_b m_b}}\right\}. \tag{28}$$

Observe that $\hat{\mu}_i^-=\mu_b^0(\pi_b)$ if $\kappa_p<2\pi_i c\kappa_b\lambda_i/(\pi_b m_b)$.

Convexity implies that $\bar{\vartheta}(\mu_t)<0$ for $\mu_t\in\left(\hat{\mu}_i^-,\hat{\mu}_i^+\right)$. Since the first case implies that there is no investment for $\mu_t>\mu_b^0(1)$, this implies that delegation produces a positive payoff only if $\mu_t<\hat{\mu}_i^-$.

Third, when optimal investment is positive and no investment results in an unviable program ($\mu_t<\underline{\mu}_i/\delta$), substituting into the analog of Equation 18 produces the corner net benefit of delegation:

$$\underline{\vartheta}(\mu_t)=\pi_i\left(\frac{\lambda_i^2-k\Lambda^2}{2\Lambda}-\lambda_i c\sqrt{\frac{\kappa_b}{\delta\pi_b m_b}}\right)+\frac{\kappa_p}{2}\left(\mu_t-\frac{\Lambda}{\delta}-\sqrt{\frac{\pi_b m_b}{\delta\Lambda_b}}\right). \tag{29}$$

Observe that $\bar{\vartheta}(\mu_t) = \underline{\vartheta}(\mu_t)$ is satisfied uniquely at $\mu_t = \underline{\mu}_i/\delta = \frac{\Lambda}{\delta} + \frac{2c\lambda_i\Lambda}{\delta\left(\lambda_i^2 - k\Lambda^2\right)}$; i.e., the capacity level such that without investment, the politician becomes indifferent between shutting down and continuing the program at $t+1$.

Combining cases, we have the group i politician's expected gain from delegation for any given μ_t:

$$\begin{cases} 0 & \text{if } \mu_t \geq \mu_b^0(\pi_b) \\ \underline{\vartheta}(\mu_t) & \text{if } \mu_t < \mu_b^0(\pi_b), \mu_t \leq \underline{\mu}_i/\delta \\ \bar{\vartheta}(\mu_t) & \text{if } \mu_t < \mu_b^0(\pi_b), \mu_t > \underline{\mu}_i/\delta. \end{cases}$$

As $\underline{\vartheta}(\mu_t)$ is increasing and linear, and $\bar{\vartheta}(\mu_t)$ is decreasing and positive for $\mu_t \in \left(\Lambda/\delta, \hat{\mu}_i^-\right)$, when $\underline{\mu}_i/\delta < \mu_b^0(\pi_b)$ the expected gain from delegation is positive for some μ_t if and only if $\underline{\vartheta}\left(\underline{\mu}_i/\delta\right) > 0$. The possibility that $\underline{\mu}_i/\delta \geq \mu_b^0(\pi_b)$ is ruled out by assumption in Equation 6.

When $\underline{\vartheta}\left(\underline{\mu}_i/\delta\right) > 0$, the monotonicity of $\underline{\vartheta}(\mu_t)$ and $\bar{\vartheta}(\mu_t)$ on $\left(\Lambda/\delta, \hat{\mu}_i^-\right)$ further imply that delegation can only occur within a convex interval over μ_t. The supremum of the set of μ_t for which the delegation gain is positive is $\hat{\mu}_i^-$. The infimum is characterized by $\underline{\vartheta}(\mu_t) = 0$. This produces the following critical value of μ_t:

$$\tilde{\mu}_i \equiv \mu_b^0(\pi_b) + \frac{\pi_i}{\kappa_p}\left(2c\lambda_i\sqrt{\frac{\kappa_b}{\delta\pi_b m_b}} - \frac{\lambda_i^2 - k\Lambda^2}{\Lambda}\right). \tag{30}$$

Thus when the region

$$\mathcal{D}_i \equiv \left(\tilde{\mu}_i, \hat{\mu}_i^-\right) \tag{31}$$

is nonempty, delegation is optimal for a group i politician. ■