16 The evolution of governance structures in a polycentric system

Edward McPhail and Vlad Tarko

INTRODUCTION

A polycentric governance system is a system of several independent centers of authority and decision-making operating under an over-arching set of formal and informal rules (Ostrom et al. 1961; Ostrom 1991b, ch. 9, 1999; Ostrom 2005, ch. 9, 2010; Wagner 2005; McGinnis and Ostrom 2012; Aligica and Tarko 2012, 2013; Tarko 2015; McGinnis 2016). The overarching set of rules is necessary because the decisions of one center create positive and/or negative externalities upon others. Despite these externalities, the centers do not merge into a single unified and centralized decision-making unit for several reasons.

First, the centers operate under heterogeneous beliefs and preferences which makes consensus-building too difficult (McGinnis 2016). In the language of the calculus of consent (Buchanan and Tullock 1962 [1999]), both the decision-making costs of centralized decision-making and the external costs of each central decision, imposed upon those who would disagree with the centralized decisions, would be too large.

Second, the decision centers address a wide range of problems, and the solutions to each of these problems can have very different optimal scales (Ostrom et al. 1961). This means that decision centers have to find a way to face the problem that the scales of operation of administrative units are rigid, while problems are fluid and come at varied, and changing, scales. As Ostrom et al. (1961) first noted, and as Elinor Ostrom and her collaborators later documented across a wide range of examples (Bish 1971; Bish and Kirk 1974; Ostrom 1976; Ostrom et al. 1978; Bish and Ostrom 1979; Ostrom et al. 1988; McGinnis 1999), the solution to this administrative rigidity problem is to have smaller administrative units cooperate on a quasi-ad hoc basis to address larger-scale problems as they appear. The side effect of this solution is that, rather than having a hierarchical public administration organization, we are left, by necessity, with a polycentric one. Different decision-centers are constantly engaged in mutual adjustment, both in terms of competing with one another and in terms of cooperating to solve larger-scale problems (McGinnis 2016). Furthermore, in line with their heterogeneity of beliefs and preferences, the cooperation is conditional, involving a certain degree of free entry and free exit.

A polycentric system usually does not have purely deliberate social-economic outcomes. Similar to a market, the outcome of the operation of a polycentric system is an emergent order which has certain unplanned features. These features may be desirable or less desirable. The way in which a polycentric system addresses undesirable outcomes is by altering the set of overarching rules, such that the emergent outcome will be improved, for example, by limiting some negative externalities or by making broader use of the knowledge discovered by one decision-center. Figure 16.1, adapted from the frameworks proposed by Aligica and Tarko (2012) and McGinnis (2016), illustrates the structural

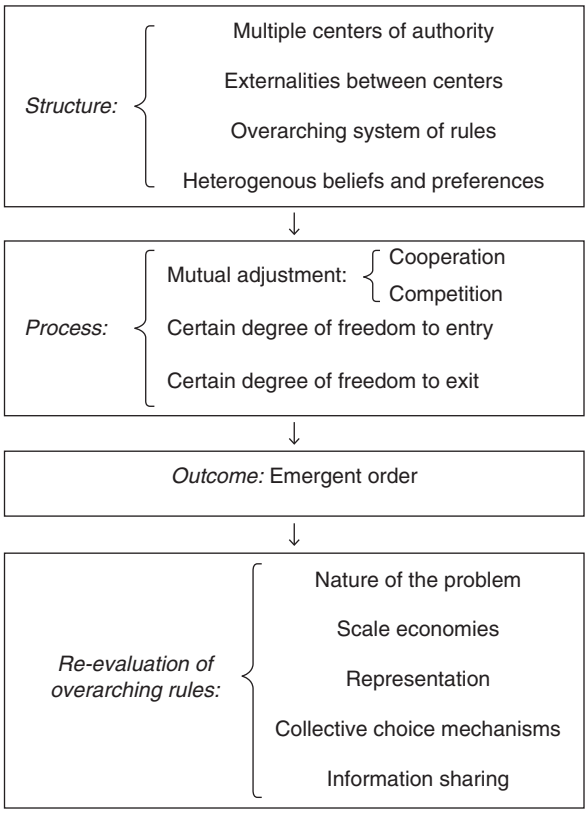

Figure 16.1 The structure and operation of a polycentric system

features of a polycentric system and the two processes involved: (1) the operational-level process, concerned with the actual production of public goods and the emergent system-level order; and (2) the collective-choice process, concerned with identifying problems, giving voice to different points of view about each problem, and reforming the overarching set of rules.

The concept of polycentricity involves a generalization into the realm of public economics of the concept of markets. Markets are one special case of polycentricity (Ostrom 1991, pp. 229–31), but polycentricity aims to capture the characteristics of productive and sustainable emergent orders even outside the operation of the price system. Other examples of polycentric systems include the scientific community, competitive local public economies, common law, and international relations (Ostrom 1991, chp. 9). Generaly speaking, the purpose of this concept is to provide insight into the conditions under which non-market emergent orders can be expected to lead to desirable outcomes. The concept of polycentricity was first developed by analogy to markets in the context of the metropolitan governance debate, when the mainstream of the public administration profession was arguing in favor of consolidating the metropolitan administrations into large centralized bodies that would take advantage of economies of scale (McGinnis

1999; Aligica and Boettke 2009; Aligica and Tarko 2012; Boettke et al. 2013, 2016). The Ostroms dissented from this intuition, thanks to an analogy to markets, and highlighted the potential efficiency of local public economies:

> Duplication of functions is assumed to be wasteful and inefficient. Presumably efficiency can be increased by eliminating 'duplication of services' and 'overlapping jurisdictions.' Yet we know that efficiency can be realized in a market economy only if multiple firms serve the same market. Overlapping service areas and duplicate facilities are necessary conditions for the maintenance of competition in a market economy. Can we expect similar forces to operate in a public economy? (Ostrom and Ostrom 1977 [1991], pp. 163–97)

From their perspective, the most likely path to efficient public administration was not consolidation, but developing smart overarching rules that would allow productive 'interorganizational arrangements'. Such arrangements 'would manifest market-like characteristics and display both efficiency-inducing and error-correcting behavior. Coordination in the public sector need not, in those circumstances, rely exclusively upon bureaucratic command structures controlled by chief executives. Instead, the structure of interorganizational arrangements may create important economic opportunities and evoke self-regulating tendencies' (Ostrom and Ostrom 1977 [1991], pp. 163–97).

From this conceptual starting point, focused on facilitating the emergence of productive bottom-up orders in the realm of public economies, the Ostroms and their colleagues have gathered comprehensive empirical evidence regarding the advantages, in terms of efficiency, voice, and resilience, of such competitive public structures (Bish 1971; Bish and Kirk 1974; Ostrom 1976; Ostrom et al. 1978; Bish and Ostrom 1979; Ostrom et al. 1988; McGinnis 1999). Half a century later, we have the empirical evidence regarding the validity of the polycentric perspective, but a new need is felt for a deeper theoretical understanding of the competitive aspect of polycentricity. The Bloomington studies of local economies have shifted the debate towards much greater skepticism regarding centralization, while supporting the idea of institutional competition. However today, the metropolitan debate, with respect to police services in particular, has reignited, owing to the failure of 'community policing' (Boettke et al. 2013; 2016), and the idea that institutional competition leads to a 'race to the bottom' still persists (for example, Geradin and McCahery 2004).

The simplest theoretical approach has been to import into public economics the model of perfect competition. Indeed, the Tiebout 'voting with your feet' model (Tiebout 1956; Ostrom et al. 1961; Donahue 1997; Caplan 2001; Howell-Moroney 2008; Boettke and Marciano 2016) is thought to generate efficient outcomes primarily thanks to an assumption of relatively low exit costs. However, as noted by authors like Buchanan and Goetz (1972), Donahue (1997) and Caplan (2001), such an assumption is often unwarranted (see also Boettke and Marciano 2016). For example, capital is often difficult to relocate. Moreover, public services and regulations are bundled into large packages and, hence, we can rarely choose the prefered bundle of public services in the same manner as we choose a prefered bundle of private goods. Furthermore, the payment for public services, in terms of taxes, also does not occur in the same straightforward manner as in private markets. Paying taxes is more similar to providing someone a grant in the hope of then receiving a particular bundle of services. To make matters worse, the rational ignorance of voters further diminishes their ability to constrain the public administration (Lyons

and Lowery 1989; Lowery 1998; Boettke et al. 2011). This makes it difficult to reveal accurately citizens' preferences about public goods, and to establish optimal levels of expenditure. As noted by Vincent and Elinor Ostrom (1977 [1991], pp. 163–97), while '[a]n expression of demand in a market system always includes reference to what is forgone as well as what is purchased', by contrast, '[t]he articulation of preferences in the public sector often fails to take account of forgone opportunities'. As they put it, '[w]hereas the income received for providing a private good conveys information about the demand for that good, payment of taxes under threat of coercion indicated only that taxpayers prefer paying taxes to going to jail. Little or no information is revealed about user preferences for goods procured with tax-supported expenditures'.

These limitations do not imply that institutional competition is meaningless; they only lead us to the conclusion that we need to model competitive public economies using oligopoly and monopolistic competition models, rather than the perfect competition model. Depending on which model of oligopoly we use (Kreps 1990, ch. 10), such as Cournot, Stackelberg or Bertrand, and focusing on the quality of public goods rather than on quantity, we would get different predictions with respect to the 'price' (that is, tax rates) and quality of the services provided. With Cournot competition, when two (or few) competing providers simultaneously choose how much to provide, we obtain a higher price than the perfectly competitive model, but lower than the monopoly price. In the case of Stackelberg competition, when one of the providers acts as a leader and the other as a follower, the result is that the leader gets a higher market share. If capital costs are low, the price under Stackelberg competition is lower than under Cournot, while the opposite holds if capital costs are high. Under Bertrand competition, the providers compete in terms of price rather than quality. If each jurisdiction could in principle satisfy the entire demand (that is, no constraints of potential capacity), the outcome of Bertrand competition, even with just two providers (as long as collusion is avoided), is the same as under perfect competition. In case of capacity constraints, some of the customers end up paying more, while the rest of the customers pay the lowest price.

It seems to us that most public economies are better described by the Cournot and Stackelberg models because of relatively weak constraints on tax collection – the public sector operates under a 'soft budget constraint' (Kornai 1986). Even so, tax competition among jurisdictions introduces a certain element of Bertrand competition into the picture. To the extent that Bertrand competition is present in public economies, it is usually under capacity constraints. Such capacity constraints are partly natural, as a result of people's preferences against too high population densities, and partly artificial, owing to immigration restrictions and constraints created by zoning laws (for example, leading to higher housing prices and rents).

This brief discussion of oligopoly models shows that, although imperfect Tiebout competition cannot be expected to lead to perfectly efficient results, we still, nonetheless, have reasons to believe that a certain tendency towards efficiency persists. In what follows, we build a model of (1) interjurisdictional competition and (2) interjurisdictional cooperation for providing larger-scale public goods, under very stringent assumptions about knowledge and benevolence. We assume that households have no knowledge about other jurisdictions, and decide purely on the basis of their satisfaction with their current jurisdiction (that is, leaving decisions are blind leaps into the unknown), and we assume that local governments operate without any inherent concern for the public interest, that

is, as revenue-maximizing Leviathans. We can see this as a 'robust political economy' (Levy 2002; Farrant 2004; Leeson and Subrick 2006; Pennington 2011; Boettke and Leeson 2012) model of the operation of a polycentric governance system. In line with the Bloomington approach, we build a model of both interjurisdictional competition and interjurisdictional cooperation. This allows us to model the bottom-up emergence and evolution of governance structures, in line with the perspective first put forward by Ostrom et al. (1961) and later elaborated both theoretically and empirically by Elinor Ostrom and others. Our robust political economy assumptions about limited knowledge and benevolence allow us to avoid some of the main sources of skepticism regarding the original Tiebout model.

The next section starts by building a simple model which assumes a given scale of jurisdictions. Section 3 relaxes this assumption describing a process by which the scale at which a public service is provided increases or decreases, hence describing the emergence and evolution of larger-scale governance structures. Section 4 shows that alternative mechanisms to exit, such as voice, emerge as a consequence of some households either not having the resources to move or still holding hope that the quality of public services will improve. We, thus, see voice as a second-best solution: only households that find it too costly to move (owing to a variety of costs ranging from simple moving costs all the way to the costs of leaving behind their social networks and social capital) use voice as an attempt to improve the public services they receive. While exit provides a direct, unconditional, benefit to the household as a result of its own action, voice provides a benefit only conditional on what other households also agree to do.

2 A MODEL OF ENDOGENOUS QUALITY JURISDICTIONS

2.1 Assumptions

Some of the classic endogenous quality (EQ) models rely on public reputation as a quality assurance mechanism (Klein and Leffler 1981; Allen 1984; Shapiro 1982, 1983; Rogerson 1983, 1987). By contrast, we develop here a model focused on individual household exit. A public reputation mechanism assumes that households communicate with one another and share their experiences of various jurisdictions. However, some of the strongest critiques of the Tiebout model rest on the empirical observation that people in fact have very limited such information (Lowery and Lyons 1989; Lyons and Lowery 1989; Lowery 1998; Boettke et al. 2011). In the EQ model presented below, inspired by Gintis (1989) and McPhail (1997, 2001), households know only the quality history of their own consumption. They do not know the quality histories of other households, only their own.

Furthermore, we analyze the situation for just one public service at a time, rather than for the bundle of all services simultaneously. At first, this may seem an odd choice, given that public services within a jurisdiction are a package deal, and local citizens and firms do not pay for each service individually, but only pay a unified tax. However, there are two strong reasons for developing the model this way.

First, as long argued by the Bloomington school, from the classic Ostrom et al. (1961) paper onward, different services are optimally provided at different scales, and the administrative structure of governance has to adapt in various ways to this diversity. In this

chapter we are not interested in actually modeling the details of the negotiation process between jurisdictions and the exact political mechanisms by which a given service ends up more or less centralized. Instead, we provide only an equilibrium model highlighting the broad presumed tendency towards which the complex underlining negotiation process tends to lead. However, the point still holds that we need to analyze the situation issue by issue.

The second reason is related to the issue of citizens' limited knowledge. Boettke et al. (2011) cite several empirical studies demonstrating that citizens have very poor comparative knowledge about the public services offered by different jurisdictions and about the tax rates in different jurisdictions. This evidence provides a strong reason to dismiss Tiebout competition models that bundle all the services together, that is, which assume that households do a comprehensive comparative analysis of all services across all jurisdictions before moving. By contrast, we assume the opposite idealization, namely, that the moving decision of a given household depends only on one service, that is, the service they regard the most important. For example, a household that cares primarily about the quality of the public schools will probably be informed about the public schools, but not about many other public services. As such, the results of the surveys mentioned by Boettke et al. (2011) are neither surprising nor relevant. Because of this more realistic assumption about knowledge, we address the Tiebout competition with respect to each public service separately. Each public service will have a different set of 'marginal citizens' who are sufficiently unsatisfied about the quality history of the public service they have received to explore their moving options. Each such set of 'marginal citizens' is much smaller than the entire population of a jurisdiction.

As this approach is based on 'marginal citizens', the model may not apply to all public services, because the least important public services may not have any marginal citizens. That is, perhaps no one's decision to move would rest on the quality of some of these least important services. As such, our model only covers the most important issues. The model can be extended by assuming that individual households' decision to move depends on more than just a single issue, but for simplicity's sake we do not cover such extensions here.

2.2 The Model

Suppose a jurisdiction produces an indivisible public good. Households enjoy one unit of the good per period. For the sake of expositional clarity, we assume that households behave according to a specific functional form. Production occurs under constant returns to scale at a per unit cost of $c(q; w)$, where q denotes the quality of the good, and $\mathbf{w}$ is the vector of factor prices. We also assume that producing a higher quality service is more expensive, that is, $\partial c/\partial q$, $\partial^2 c/\partial q^2 > 0$ for $q > 0$, $c(0) = \partial c/\partial q(0) = 0$, and that there is a finite q_{max} such that $\lim_{q \to q_{max}} c(q) = \infty$. Let $T > 0$ be the tax or fee paid by the household to the jurisdiction for the good, and we denote jurisdiction's per unit profit (rent) as $\pi = T - c(q; \mathbf{w})$. In practice, a jurisdiction collects taxes, and then has a separate fiscal decision-making process, determining how much to allocate to fund the production of each public good. For our purposes, it is sufficient to note that a household pays a de facto sum T for a given service. This T is an imperfect equivalent of the price citizens' pay for the service, although, as we have mentioned in the introduction, this 'price' is unlikely to properly reflect the opportunity costs of providing a particular quality of the public good.

We define the endogenous quality model as consisting of households that behave according to the moving-out algorithm described below and of jurisdictions that behave as revenue-maximizing Leviathans.

2.2.1 Households moving algorithm

In what follows we focus on households 'voting with their feet'. However, the exact same logic applies to the movement of capital; indeed, some of the criticism of the Tiebout model has focused on capital more than on households (for example, Caplan 2001). Although, for simplicity, we focus on households, the results should, thus, be interpreted in a more general fashion as referring to both households and firms.

We consider a set of households who stay with a jurisdiction for a number of periods, and then contemplate moving when dissatisfied. These are the 'marginal households' for the public good under consideration, that is, those households who care about this public good above all others and whose moving decision is influenced by the quality of this public good. We assume that each household chooses a level of service quality q_{crit} and treats the service in each period as a success if, and only if, the observed quality is greater than q_{crit}. As mentioned, the critiques of Tiebout competition focus mainly on the existence of prohibitive moving costs. In our model, moving costs simply imply that a higher rate of failure is required before the household decides to move, that is, moving costs cause a decline of q_{crit}. The household aspires to attain a certain quality level, q_{asp}, and faces moving costs, c_{move}, so that $q_{crit} = q_{asp} - c_{move}$ where q_{asp}, c_{move} Î [0,1]. The moving costs may include a wide variety of factors, from simple transportation costs and search costs, to the psychological costs due to severing local social connections. If $c_{move} = 1$ then even for very high aspiration quality such as $q_{asp} = 1$, the household would find the costs of moving to be prohibitively high and would never move. We assume that when $q_{crit} \leq 0$, the district still provides some minimum level of quality $q_{min} > 0$.

A household leaves the jurisdiction when the average number of failures exceeds the average number of successes. Feller (1968) demonstrates that this behavior follows a Bernoulli random walk, which becomes relevant in the jurisdictions' calculus below.

Furthermore, in line with the empirical evidence that people have limited knowledge about the quality of services and the tax rates in other jurisdictions, we make the weakest possible assumption about household information: when a household decides to move from one jurisdiction, it chooses another jurisdiction at random. That is, for people who have very poor information about the quality of services and tax rates in other jurisdictions, it is as if they are randomly choosing where to move.

2.2.2 Jurisdictions as revenue-maximizing Leviathans

Let $\Phi_t(q)$ be the probability that a household remains with a jurisdiction supplying quality q for t periods. If $\rho > 0$ denotes the discount rate for the jurisdiction, so that

$$\delta = \frac{1}{1+\rho} \in (0,1) \tag{16.1}$$

denotes the discount factor, then the present value of profits from a household who stays with the jurisdiction for exactly t periods is given by the following expression:

$$\pi(\delta + \delta^2 + ... + \delta^t) = \pi\frac{1 - \delta^t}{\rho} \tag{16.2}$$

The expected profit stream from a household that moves out after a finite number of periods t is

$$\sum_{t=1}^{\infty}\pi\frac{1 - \delta^t}{\rho}\phi_t \tag{16.3}$$

while that from a household that never switches is

$$\frac{\pi}{\rho}\left(1 - \sum_{t=1}^{\infty}\phi_t\right) \tag{16.4}$$

Thus, total expected profit is the sum of the two

$$E[\pi] = \sum_{t=1}^{\infty}\pi\frac{1 - \delta^t}{\rho}\phi_t + \frac{\pi}{\rho}\left(1 - \sum_{t=1}^{\infty}\phi_t\right) = \frac{\pi}{\rho}\left(1 - \sum_{t=1}^{\infty}\delta^t\phi_t\right) \tag{16.5}$$

In accordance with the revenue-maximizing Leviathan concept (Brennan and Buchanan 1977; Buchanan and Brennan 1980, 1985; Engineer 1990; McGuire and Olson 1996), we assume that each jurisdiction is trying to maximize its profits by collecting as much taxes as possible and providing the cheapest possible services. However, even under such a harsh assumption, the possibility of exit provides a check on exploitation. If there is an interior solution for q, maximum expected profits occur when q satisfies the jurisdiction's first-order condition:

$$\rho\frac{dE[\pi]}{dq} = -(T - c(q;\mathbf{w}))\sum_{t=1}^{\infty}\delta^t\frac{\partial\phi_t}{\partial q} - \frac{\partial c}{\partial q}\left(1 - \sum_{t=1}^{\infty}\delta^t\phi_t\right) = 0 \tag{16.6}$$

Based on equations (16.5) and (16.6) we can now derive a series of consequences.

2.3 Consequences of the EQ Model

Theorem 1: Every positive tax yields an interior jurisdiction optimum at which the tax exceeds marginal cost and the jurisdiction earns a positive rent per unit.

Proof: Rearranging equation (16.6) and solving for T gives:

$$T = c(q,\mathbf{w}) - \frac{\partial c}{\partial q}\left(1 - \sum_{t=1}^{\infty}\delta^t\phi_t\right)\left(\sum_{t=1}^{\infty}\delta^t\frac{\partial\phi_t}{\partial q}\right)^{-1} \tag{16.7}$$

To sign the rent note that $\partial c/\partial q > 0$ and that $(1 - \sum_{t=1}^{\infty}\delta^t\phi_t) > 0$. Hence for

$$\frac{\partial c}{\partial q}\left(1 - \sum_{t=1}^{\infty}\delta^t\phi_t\right)\left(\sum_{t=1}^{\infty}\delta^t\frac{\partial\phi_t}{\partial q}\right)^{-1} > 0, \tag{16.8}$$

we need to show that

$$\frac{\partial \phi_t}{\partial q} < 0. \tag{16.9}$$

Recall that q denotes the probability that a unit of the good is a success. By Feller (1968) the number of successes over failures follows a Bernoulli random walk where the transition probabilities are $(1 - q)$ up and q down. Feller shows that the probability of stopping at period t is the probability of a first zero crossing of a random walk at period t, which has a generating function satisfying the following quadratic equation:

$$\phi(\delta) = (1 - q)\delta + q\delta\phi^2(\delta) \tag{16.10}$$

Additionally, Feller provides the solution for $\phi(\delta)$, giving the following:

$$\phi(\delta) = (1 - \sqrt{1 - 4q(1 - q)\delta^2})/2q\delta \tag{16.11}$$

We calculate that:

$$\frac{\partial \phi}{\partial q} = -\frac{\delta(1 - \phi^2)}{1 - 2q\delta\phi} \tag{16.12}$$

To show $\frac{\partial \phi_t}{\partial q} < 0$, we need only show that $2q\delta\phi < 1$ for $\delta, q \in (0,1)$. However, $2q\delta\phi \geq 1$ implies that $2q\delta\phi^2 \geq$ f. In turn this and $\phi(\delta) = (1 - q)\delta + q\delta\phi^2(\delta)$ implies that $2(\phi - (1 - q)\delta) > \phi$, which gives $\phi > 2(1 - q)\delta$. Since $q \geq \frac{1}{2}$ in this case, this implies $\phi > \delta$. But this contradicts $\delta - \phi = (1 - 2\phi)q\delta \geq 0$. Hence the rent is positive and $T \geq c(q; \mathbf{w})$.

Discussion of theorem 1: In equilibrium, the tax equals marginal cost plus a rent. Even with zero moving costs, the rent is still positive. The limited knowledge of households ensures this outcome. The rent is the ratio of the expected marginal cost of quality incurred over a household's tenure with the jurisdiction, and the discounted marginal change in the length of stay owing to a change in quality. It represents the tradeoff the jurisdiction faces as it raises quality. Increasing quality means that, on the one hand, costs rise per unit of output, but on the other hand, the expected length of time a household stays in a jurisdiction rises as well. Because of our assumption of zero-interjurisdictional knowledge on the part of the households, jurisdictions cannot attract more households by raising quality, they can only retain them longer.

It is instructive to rearrange equation (16.6) as follows:

$$\underbrace{(T - c(q;\mathbf{w}))}_{\pi}\left(\sum_{t=1}^{\infty}\delta^t\frac{\partial \phi_t}{\partial q}\right) = -\frac{\partial c}{\partial q}\left(1 - \sum_{t=1}^{\infty}\delta^t\phi_t\right) \tag{16.13}$$

The left-hand side is the discounted expected marginal rental stream from one household due to an increase in quality. The right-hand side is the expected discounted marginal cost

of quality for one household. Thus, the jurisdiction chooses quality such that the marginal cost of quality equals the marginal profit from keeping more taxpaying citizens,

$$\frac{\partial c}{\partial q} = \underbrace{(T - c(q;\mathbf{w}))}_{\pi}\underbrace{\left(- \sum_{t=1}^{\infty}\delta^t\frac{\partial\phi_t}{\partial q}\right)\left(1 - \sum_{t=1}^{\infty}\delta^t\phi_t\right)}_{\eta(q)} \tag{16.14}$$

or,

$$\frac{\partial c}{\partial q} = \pi(T,q,\mathbf{w})\,\eta(q) \tag{16.15}$$

The right-hand side is the additional profit the jurisdiction receives from an increase in quality owing to the increased expected length of stay by a given household. The left-hand side is the additional cost incurred by the jurisdiction from this increase in quality. Here, $\eta(q)$ denotes the quality elasticity of sales. It shows how responsive the length of stay of a household is to changes in quality. Thus, it tells us how many more citizens are retained by the jurisdiction and therefore how much the demand for public services rises when quality rises. As mentioned, these citizens are from the 'marginal group' with respect to the public service under analysis, that is, citizens who primarily care about this service.

Theorem 2: Quality is an increasing function of the tax, and a decreasing function of factor prices.

Proof: Taking the total differential of equation (16.6), treating ρ as a constant, and q and T as variables, we find that

$$\rho\frac{\partial^2 E[\pi]}{\partial q^2}dq - \rho\frac{\partial}{\partial q}\left(\sum_{t=1}^{\infty}\delta^t\phi_t\right)dT = 0 \tag{16.16}$$

and solving for dq/dT gives

$$\frac{dq}{dT} = \frac{\frac{\partial}{\partial q}\left(\sum_{t=1}^{\infty}\delta^t\phi_t\right)}{\frac{\partial^2 E[\pi]}{\partial q^2}} > 0 \tag{16.17}$$

by the firm's second order condition and the fact that

$$\frac{\partial}{\partial q}\left(\sum_{t=1}^{\infty}\delta^t\phi_t\right) < 0, \tag{16.18}$$

as shown in the proof to theorem 1.

Costs are a function of factor prices, $\mathbf{w}$, as well as quality, $c(q;\mathbf{w})$. Differentiating equation (16.6) with respect to factor price k, w_k, yields

$$\frac{\partial q}{\partial w_k} = -\frac{1}{\dfrac{\partial q(w,T)}{\partial T}}\left(\frac{\partial c}{\partial w_k} + \frac{1}{\eta(q(w,T))}\frac{\partial^2 c}{\partial q\,\partial w_k}\right) < 0 \tag{16.19}$$

Discussion of theorem 2: Despite the fact that we are using the ultra-pessimistic revenue-maximizing Leviathan concept, and despite the minimalistic assumptions about households' knowledge, we still obtain the conclusion that jurisdictions are forced to provide an increased benefit when raising taxes.

These results also show that if we know the tax rate, or more precisely, the fraction from the taxes collected in the jurisdiction that go to paying for the public service we are analyzing, and the factor prices, the model determines the quality of the service. This is why we refer to this model as an endogenous quality model. Different jurisdictions may differ either in terms of their implicit tax rates for each service (which is a combination of the overall level of taxes and budget allocation decisions), or in terms of the factor prices. Geographical and other differences between jurisdictions are captured by these factor prices. Various federal-level policies, such as subsidizing land development in certain areas or establishing differential policy regimes in different geographical areas, are also included here in the factor prices.

As we would expect, if factors of production become more expensive, the quality of services using those factors declines – unless a corresponding increase in taxes occurs. This can have an interesting dynamic in terms of local economic development. For example, public services such as quality roads contribute to development by reducing transaction costs. If labor costs increase, this will have a negative effect on road maintenance. The decrease in road quality could be prevented by an increase in taxes; however, increasing taxes itself can have a negative impact on development. So, shocks to factor prices can have wide-reaching consequences via a chain of effects.

Similarly, the model readily accommodates shocks to demand. Consider an especially large outflow of households from a jurisdiction. As a result, the remaining households are more valuable to the local government. We model this via a decrease in ρ or, equivalently, an increase in δ.

Theorem 3: For a given tax, a sudden outflow of households will lead, ceteris paribus, to a rise in quality and a fall in the jurisdictional rent.

Proof: The outflow of households is modeled as an increase in δ. Taking the total differential of equation (16.2), noting that $\rho = (1 - \delta)/\delta$, and treating q and δ as variables, we find that

$$\rho\frac{\partial^2 E[\pi]}{\partial q^2}dq + \left(\rho\frac{\partial^2 E[\pi]}{\partial q\,\partial\delta} - \frac{1}{\delta^2}\frac{\partial E[\pi]}{\partial q}\right)d\delta = 0 \tag{16.20}$$

and solving for $dq/d\delta$ gives

$$\frac{dq}{d\delta} = -\frac{\rho\frac{\partial^2 E[\pi]}{\partial q \partial \delta} - \frac{1}{\delta^2}\frac{\partial E[\pi]}{\partial q}}{\rho\frac{\partial^2 E[\pi]}{\partial q^2}} > 0 \tag{16.21}$$

where

$$\frac{\partial^2 E[\pi]}{\partial q \partial \delta} > 0, \frac{\partial E[\pi]}{\partial q} = 0 \tag{16.22}$$

by the first order condition, and

$$\frac{\partial^2 E[\pi]}{\partial q^2} < 0 \tag{16.23}$$

by the second order condition.

The rent, $T - c(q; w)$, decreases because, as $\partial c/\partial q > 0$, a higher quality service implies a higher cost.

Discussion of theorem 3: Consider a jurisdiction such as Detroit which has suffered large outflows of households. One response is for the city to attempt to provide greater quality since each household relationship is now more valuable but, as in the case of labor cost shocks that we discussed above, this 'negative demand shock' has additional effects that can undermine the attempt to raise quality. Urban flight may well undermine the web of social connections that made old Detroit a more attractive place to live, so, as the number of households falls, the social capital is also destroyed, further reducing the value of remaining in Detroit.

Interestingly, such considerations about 'social capital' are already implicitly included in our model thanks to the mathematical properties of the Bernoulli random walk. We, thus, do not need to include them as special, additional factors. The Bernoulli random walk stipulates that the longer a household resides in a jurisdiction, the more successes it has, and, therefore, the more failures it will take to cause the household to contemplate moving. We can interpret this as a kind of reservoir of goodwill that the jurisdiction has built up. The greater the number of successes, the greater the length of time a household resides in a jurisdiction, and the more 'forgiving' the household is bound to be, in the sense that it will take more failures to fully erode that goodwill. Recall that the household considers switching only when the average number of failures exceeds the average number of successes. So, for a household that has resided in a jurisdiction for t periods, as t grows large, the number of failures required to prompt them to consider switching grows as well. This stickiness or reluctance to move comes not just from the successes themselves but also from the duration of time. We may imagine that the longer the time a household resides in a jurisdiction the more rooted it becomes; the greater the connections – social and economic – it has made. The Bernoulli random walk catches the flavor of this intuition, although we have not included an explicit mechanism in our model deliberately for this purpose. In section 4 we consider further elaborations of this idea based on the concepts of co-production and voice. The EQ model developed so far does not allow for declines in quality to coexist with declines in population; but, empirically, such things arguably occur.

It is the concept of co-production, introduced in section 4, that allows us to account for such situations.

While more work remains to be done investigating the complex interaction of these effects, our sketch above describes some of the intuition provided by the EQ model. If critical quality is a function of the time a household has spent in a jurisdiction, then more structure is needed. With a more fully specified treatment we can investigate how age composition of households, income differences, preference heterogeneity, and others have on the provision of quality.

3 THE SCALE OF PUBLIC SERVICES

The previous section has endogenized the quality of the public services, but has not addressed the issue of scale. Different services are best provided at different scales. To account for the scale at which the public service is provided, we start by considering the smallest, most local, jurisdictions as our basic unit of analysis.

For example, in Buchanan's (1987) social contract account, the basic units are individuals. He noted that we should not assume 'without inquiry, that the individual [is] locked into membership in a political community and that the range and the scope of the collective's activities [are] beyond the control of the individual' (Buchanan 1987, p. 306). We need to understand 'the conditions that must be present for the individual to find it advantageous to enter into a political entity with constitutionally delineated ranges of activity or to acquiesce in membership in a historically existent polity' (Buchanan 1987, p. 306).

In our account, we use the smallest local jurisdictions as the basic institutional unit, rather than individuals, but Buchanan's point still stands, that we need to have a mechanism by which the scale at which a public service is provided increases or decreases. The previous section simply assumed that the scale at which a public service is provided is given, and focused on endogenizing the quality of services and the rents each jurisdiction provides. We now endogenize scale as well. We define an endogenous quality polycentric system as consisting of EQ jurisdictions and households, and having a structure emerging by the process described in this section.

3.1 Assumptions

We assume that larger scale provision results from smaller scale jurisdictions collaborating to create a larger scale organization for the provision of particular services which will be available to households in all the participating jurisdictions. We refer to the number, n, of these smallest local jurisdictions, combining into a larger administrative unit to provide the public service that we are analyzing. This number measures the degree to which services are centralized. If the service is completely centralized, $n = 100$ percent, that is, all local jurisdictions have merged into a single collaborative unit, while lower n indicate various levels of decentralization.

This way of thinking about scale fits a wide range of historical examples of both increased centralization and increased decentralization. For example, the formation of the United States or of the European Union (EU) involved a negotiation process among smaller administrative units for the creation of a larger-scale association. In line with our

model that uses jurisdictions as the unit of institutional analysis, such examples involved the association of pre-existing jurisdictions, rather than of individuals. Similarly, examples of associations breaking up, such as the collapse of the Western Roman Empire, the gradual reduction of the Byzantine Empire, or more recently, the break-up of the Soviet Union or the exit of Singapore from Malaysia, also involve local jurisdictions exiting previous associations.

The same phenomenon also happens at much more local levels. Consider a typical example of common pool resources studied by Elinor Ostrom (1990). Following a severe overfishing issue in Maine lobster fisheries in the 1920s, the local communities were able to address the free-riding problem by appealing to a larger-scale organization, the state of Maine, which 'supported informal local enforcement efforts' (Ostrom 1999a, p. 40). More recently, the informal local fishing organizations were transformed into formalized councils with democratic local elections and formalized authority over specified geographical areas. This facilitated the cooperation between local communities in regard to large-scale problems and the bottom-up emergence of larger-scale associations: 'the formalization of local zones was followed, almost immediately, by the creation of an informal council of councils to address problems at a greater than-local scale' (Ostrom 1999a, p. 40). A similar process occurred with respect to Washington state Pacific salmon fisheries, where salmon over-fishing was solved by means of 'a 'co-management' system that involves both the state of Washington and the 21 Indian tribes in diverse policy roles related to salmon' (Ostrom 1999a, p. 40).

These examples show that cooperation between local communities sometimes rests on using pre-existing larger-scale administrative units. In our simplified model, however, we consider only the bottom-up process.

3.2 Organization Costs

In line with the above polycentric way of thinking about how larger-scale public services are provided by an emergent collaborative agreement between smaller-scale administrative units, let us define the transaction cost, $F(n)$, involved in setting up the larger-scale collaborative agreement. This cost is the same as the 'decision-making cost' from the calculus of consent model, and, as argued by Buchanan and Tullock (1962 [1999], ch. 6), $F(n)$ is an increasing function, $F' > 0$. The larger the coalition of jurisdictions, that is, the more centralized the provision of the public service, the greater the decision-making cost for setting up the association. $F(n)$ can also be understood as an entry fee that each jurisdiction pays as a cost of becoming part of the larger collaborative association. For example, in case of the EU's expansion, F is the cost of negotiating during the accession process and the costs of pre-accession reforms requested by the EU. Because of the need for consensus in accepting a new country into the union, the larger n is, the more difficult the entry of new countries becomes.

To develop a model of scale and equilibrium, more structure must be added to the cost side of our model. If the only costs that a jurisdiction faces are variable, then positive per-unit rents combined with constant returns to scale always imply positive economic profits *ex ante* entry and, consequently, an indeterminate scale. As is well known in the product quality literature, there exist various ways by which profits can be dissipated in equilibrium.[1] In our model, this is owing to the fact that as more jurisdictions join the association, the decision-making costs increase.

Apart from the joining fee F, we also assume that the jurisdiction incurs a per period fixed cost, f, which we can see as an ongoing tax for being part of the larger association. This fee accounts for things like the continued monitoring and enforcement of the rules of the larger association.

3.3 Centralization

Our model depends on jurisdictions forming expectations about how many households they are going to have, but does not depend on specific assumptions about how exactly they form these expectations. This includes anything from rational expectations, to adaptive expectations, to utterly irrational expectations. Let h_0 be the actual number of households that a jurisdiction has upon entry, and h_t as the number of households a jurisdiction has in period t. Then, for our purposes, all we need to know here is that $E_0(h_t|h_0)$ denotes the jurisdiction's expectation, formed before entry, of the number of households in period t, given the actual number of households that a jurisdiction has upon entry.

For example, the public service may have certain economies of scale or there may be positive interjurisdictional externalities that would be facilitated by the association (as in the common-pool examples given above). As such, individual jurisdictions can only provide by themselves a relatively low-quality service, and, hence, satisfy few households. Once they form the association, the quality of the services increases, each of them keeping more households (while jurisdictions that have not entered the association would lose households at a higher rate). Thus, by entering the association, the jurisdiction receives

$$E[\pi(n)] = \sum_{t=1}^{\infty} \delta^t \left[\underbrace{(T - c(q, \mathbf{w}))}_{\pi} E_0(h_t|h_0) - f \right] \tag{16.24}$$

as the expected discounted profit stream for the life of the inter-jurisdictional association.

Theorem 4: In an EQ polycentric system there exists a unique equilibrium (F^*, n^*).

Proof: If the present value of entry exceeds the entry fee, then the service becomes more centralized as more jurisdictions enter. Jurisdictions join the association as long as $E[\pi(n)] \geq F(n)$. The expansion of the association ends when

$$\sum_{t=1}^{\infty} \delta^t [\pi E_0(h_t|h_0) - f] = F(n) \tag{16.25}$$

Thus, in equilibrium, jurisdictions are indifferent between entering and staying out of the association. This provides the equilibrium level of centralization, n^*, and, implicitly, the equilibrium entry fee, $F^* = F(n^*)$.

Discussion of theorem 4: Perhaps the most surprising aspect of this model is that the entry fee is endogenized as well. That is, the jurisdictions that are already part of the association cannot require an arbitrarily large entry fee, nor can the new entrant require an arbitrarily low entry fee. What happens is that the existing jurisdictions estimate a certain increase in their individual rents as a result of accepting the new member, which

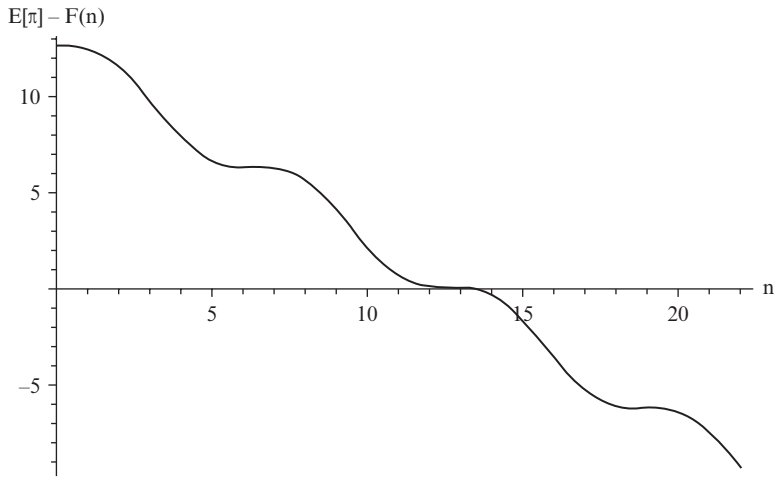

Figure 16.2 Equilibrium level of centralization

incentivizes them to keep the entry fee low. The costs of allowing a new jurisdiction in depend on whether the existing jurisdictions think they are at 'full capacity' of households already. If they are not, they want more immigrant households (such that tax revenues increase). However, if joining the association improves the public services within the new jurisdictions, fewer households would leave. This may be more than compensated by the fact that joining the association may significantly reduce the moving costs. If the existing jurisdictions think they are already at 'full capacity', they would perceive the arrival of further immigrant households as a cost, rather than as a benefit. This, for instance, seems to fit the perceptions driving the anti-EU campaign in Britain.

We can also have a graphical representation of theorem 3, by plotting the difference $E[\pi(n)] - F(n)$ as a function of n. The intersection with the horizontal zero axis provides the equilibrium scale of the public service, n^* (Figure 16.2).

3.4 Decentralization

Every period, a jurisdiction must decide whether to remain part of the association or exit (Buchanan and Faith 1987). Taking the number of households as given in period t, a jurisdiction decides to continue to be part of the association into period $t + 1$ by comparing the expected future rents from staying in to the future stream of costs. Jurisdictions exit when their expected profits are negative. Let h_0 denote now the number of households a jurisdiction has upon entry. Successive per period household numbers are denoted h_1, h_2,. . ., and so on. A jurisdiction decides to exit the association when

$$\pi \sum_{t=1}^{\infty} \delta^t E_0(h_t|h_0) < \frac{f}{\rho} \tag{16.26}$$

Note the role that f plays. As long as the jurisdiction is not yet at full capacity, and as long as there remains a positive probability of attracting another household, no matter how far off into the future this may occur, without a fixed per period cost, the jurisdiction

would never choose to exit the association. With a fixed per period cost, a jurisdiction will choose to exit the association if its current number of households gets below a critical minimum[2] or above the perceived full capacity level. If the membership fee, f, increases for some reason, than the association is in danger of breaking up. This is, for example, a common explanation of the fall of the Western Roman Empire (focused more on internal rather than external factors) – when the costs paid by different provinces of being part of the empire increased, they no longer had a vested interest in continuing to be part of the empire, and the center did not have the military capacity to force them to stay (Jones 1986; MacMullen 1990). Similarly, the recent difficulties that the EU has faced because of the refugee crisis have suddenly increased the costs of the union, leading to moves towards dissolution, such as suspending the open borders within the Schengen region (Alderman and Kanter 2016; Rankin 2016).

Consider also another different example. As we have seen earlier, according to theorem 2, if the price of factors increases, the quality of the services decreases. Consequently, a price shock to some factor markets may lead to a decline in the quality of service that the association of jurisdictions can provide, leading to a decline in the number of households. This may push some of the member jurisdictions below their critical number of households, leading them to exit the association, despite the fact that this will cause an even greater reduction in quality. Thus, revenue-maximizing Leviathans can spiral into vicious cycles.

Conversely, we can also understand the possibility of the opposite type of vicious cycle, namely, to an over-centralization process. When centralization increases, the moving costs increase, and, hence, the probability of moving out gets lower. As a result, once households moving costs are increased, further centralization may become profitable for the jurisdictions. Centralization eliminates the variety across jurisdictions and, hence, benefits the local governments providing lower quality services. That is, failing local jurisdictions, on certain margins, are more likely to call for centralization on those margins, even if this comes at the expense of higher quality jurisdictions.

3.5 Comparison with the calculus of consent optimum

The model we developed above, from the perspective of local jurisdictions that act as revenue-maximizing Leviathans, does not lead to the same conclusion as the calculus of consent optimum. Buchanan and Tullock described the centralization/decentralization optimum in the following way: 'The group should be expanded so long as the expected costs of the spillover effects from excluded jurisdictions exceed the expected incremental costs of decision-making resulting from adding the excluded jurisdictions' (Buchanan and Tullock 1962 [1999], p. 113). As mentioned previously, $F(n)$ corresponds to the calculus of consent decision-making costs. Let us denote the costs of interjurisdictional spillover externalities as $S(n)$. Mathematically, the calculus of consent optimum level of centralization is given by $ds = -dF$ or, equivalently,

$$\frac{d}{dn}(S(n) + F(n)) = 0 \tag{16.27}$$

Buchanan and Tullock (1962 [1999], pp. 44–8) refer to the total costs, $S(n) + F(n)$, as 'the costs of social interdependence'.

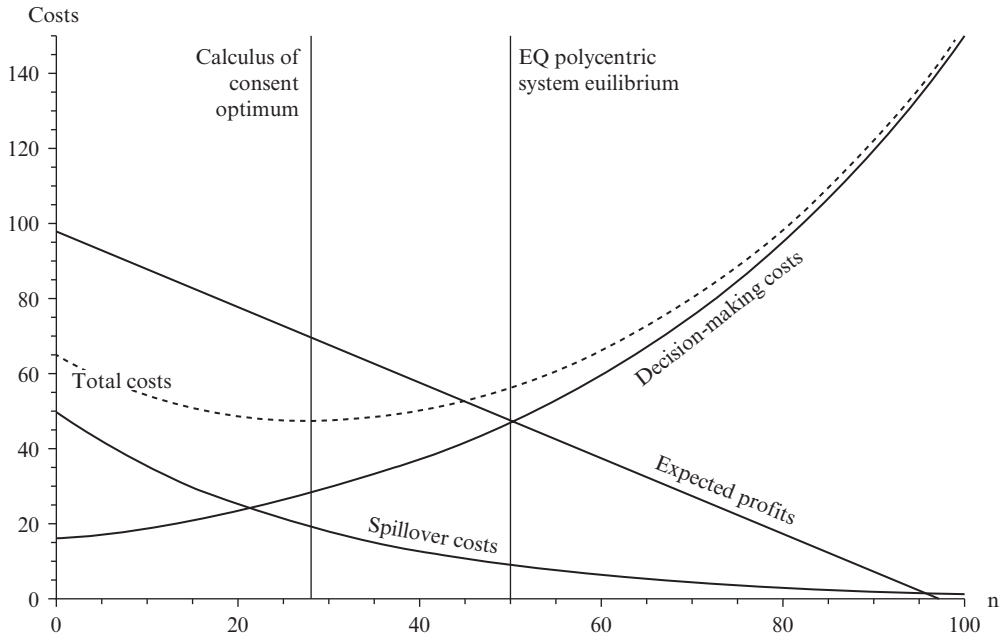

Figure 16.3 Comparison with calculus of consent optimum

By contrast, in our model, jurisdictions decide to enter or exit a larger association based not on interjurisdictional externalities but on their expectation of profits. That is, the collective association that provides the public service gradually expands up until the entry fee becomes higher than the expected profit, which may happen either before the calculus of consent is reached, in which case the polycentric system remains too decentralized, or after, in which case the polycentric system becomes too centralized (Figure 16.3).

Therefore, an important question for future research becomes: under what conditions the profit seeking mechanism of our Tiebout model converges with the calculus of consent optimum? Going back to the framework of the calculus of consent, the question is whether we can have a viable 'invisible hand' mechanism operating in the political realm in the same way as we have one in the market realm:

> Adam Smith and those associated with the movement he represented were partially successful in convincing the public at large that, within the limits of certain general rules of action, the self-seeking activities of the merchant and the moneylender tend to further the interests of everyone in the community. An acceptable theory of collective choice can perhaps do something similar in pointing the way toward *those rules for collective choice-making, the constitution, under which the activities of political tradesmen can be similarly reconciled with the interests of all members of the social group*. (Buchanan and Tullock 1962 [1999], p. 22 emphasis added)

We have no answer to this question here, but we highlight it as an important concern for future research and point out that, at least prima facie, our imperfect Tiebout competition model does not seem to be bounded towards the calculus of consent optimum. As highlighted by the above quote, this may be because we have not discussed possible

constitutional rules for constraining the Tiebout competition towards the social optimum. Our model opens the door for conceptualizing such constitutional rules more rigorously.

4 FURTHER IMPLICATIONS

Our framework is compatible with the view that exit and voice are complementary (Hirschman 1970; Oakerson and Parks 1988; Lyons and Lowery 1989), but, nonetheless, exit takes precedence. This is because, in case of exit, individual action has an immediate effect, while, in case of voice, individual action only has an effect contingent on the actions of others. Let us give a brief overview of how voice emerges as a strategy for the case when the moving condition is not satisfied – either because there is still hope that the jurisdiction will perform better or because the household does not have the resources to move.

4.1 Voice and Co-production

According to the co-production model (Parks et al. 1981; Brandsen and Pestoff 2006; Aligica and Boettke 2009; Aligica and Tarko 2013), the quality of the public service depends not only on the local government, but also on the involvement of the households. We can model this as a Cobb-Douglas production function (Aligica and Tarko 2013):

$$q = kH^{\alpha}G^{\beta} \tag{16.28}$$

where G is the contribution of the local government, and H is the average involvement of all households,

$$H = \frac{1}{N}\sum_{i=1}^{N} h_i \tag{16.29}$$

with N the total number of households in the jurisdiction and h_i the level of involvement of household i.[3]

We now have the following implication. When one household observes that $q < q_{crit}$, but the moving condition is not yet satisfied, that is, average number of failures is still smaller than the average number of successes, the household is going to increase their involvement, that is, their h_i increases causing a slight increase in H. This may lead to $q < q_{crit}$ in the next period, especially if many households engage in the same kind of action, or it may prove insufficient. If it proves insufficient for a sufficient number of periods, the average number of failures eventually gets higher than the average number of successes, and the household moves.

For example, suppose that the quality of education decreases below the critical level. As a result, if they still do not move out of the jurisdiction, parents will get more involved with education. The parents may get involved with after-school tutoring, may have fund-raisers for the school, and they may pressure the school principal. As another example, suppose that the quality of policing decreases below the critical level. The households may dedicate more time to the neighborhood watch.

As we have already seen with theorem 3, as a result of the Bernoulli distribution, a long history of past successes will make households more reluctant to leave. We can now add

the implication that it will also determine the households to engage in more co-production. This reflects a more specific mechanism for creating social capital. The households that have been in the jurisdiction for longer will have greater social capital, that is, they are more likely to engage in co-production. They act as if they have a lot more at stake in the quality of the jurisdiction.

We may further consider possible modifications of the model that incorporate the concept of loyalty (Hirschman 1970). One way would be through the effects that loyalty might have on the critical quality parameter. The longer a household resides in the jurisdiction, the lower the critical quality. This is perhaps the simplest way to capture the effect lengthening the expected stay of the household. Another way, which, we believe, better captures the intuition, would leave the critical quality unchanged. Instead, as noted above, the longer a household resides in a jurisdiction, the more likely the household is to engage in co-production and voice to try to make the quality in the jurisdiction higher than it was before. The household will desire to provide resources towards the provision of a public good or goods. The longer the household has been in the community and the greater the number of connections the household maintains with other people in the community, the more likely they are to engage in community activism and community activities. The longer a household resides in a particular jurisdiction, the greater the social network, the greater the social connections, the more it feels like home.

To return to our previous discussion of Detroit, we can imagine that the large exodus of households would adversely affect the web of social connections, making it feel less like home, reducing social capital and causing loyalty to fall. Hence, a negative demand shock is not the only adverse event. There are negative knock-on effects as well leading to an additional outflow of households that will affect co-production and voice.

4.2 Income Effects

There are two types of income effects that affect moving decisions. Recall that the critical minimal quality was defined as $q_{crit} = q_{asp} - c_{move}$, the difference between the minimum aspirational level and a factor that accounted for moving costs. On the one hand, income may lower the minimum aspirational level of quality. On the other hand, households with lower income, but high minimum aspirational quality, will be dissatisfied but will be less able to move.

Poorer households have fewer resources to contribute to the co-production of public services. For example, a single mother that has more than one job will have less time to spare to contribute to her child's education, to pay for tutoring, or to get involved to pressure the public school. As a result, poorer households may have realistic expectations of lower-quality services, which will lead them to have a lower minimum aspirational quality. As such, low-income agents are less likely to move from their current jurisdictions even if they could afford the moving costs. By contrast, high income agents would be more likely to move, further reducing the quality of services because of the decline in overall co-production participation. Consequently, those districts that initially were populated more with low-income households would tend to keep low-income residents longer and the higher income residents would tend to move out. We see that it is the co-production aspect of the public goods production that is generating this sorting effect. This is a very different mechanism from the Schelling discrimination model, and other homophily

models, which rest on preferences about neighbors (Kossinets and Watts 2009). In our case, households only have preferences about public services, but, nonetheless, can end up sorting by income.

Some low-income households may nonetheless maintain a high aspirational quality. Some may have had a very high baseline to begin with. Also, a low-income household can have a low minimum critical quality for the package of many services, but still have a preference for high quality of some of the services, for example, schooling. The fact that such low-income households may find it difficult to move may counterbalance to some extent the negative effect discussed in the previous paragraph. Unable to move away, they will tend to use voice and the ballot box as a second best option, and they will also be willing to engage in more co-production. These households will have a desire to pursue co-production, community-level service provision, voice via the political system, and other such measures whose production makes economic sense because of the lower opportunity cost of the households. Those households do not exercise the exit option, but instead they substitute into co-production and voice with respect to the services that they find wanting.

This second effect is probably relatively low for most local governance issues, which explains why lower-income jurisdictions tend to have lower-quality services, but it is a much more significant factor for large-scale issues. When centralization is very large, the moving costs are prohibitive for a very large segment of the population. While above we were assuming that only the poorest households are trapped, with highly centralized issues almost everybody is. Therefore, many will have resources to spare for using voice as an alternative strategy for trying to improve public services. We can see this as the origins of voice.

This relatively unintuitive prediction follows: we are more likely to see people engage in voice with respect to highly centralized issues than with respect to local matters. This is unintuitive because collective action is easier within smaller groups (Olson 1971). However, exit is also easier for highly local issues.

5 CONCLUSION

We have laid out an endogenous quality polycentric system model of citizens' behavior and local governments' behavior that combines:

- interjurisdictional imperfect competition, determining the quality of public goods and tax rates within each jurisdictions as a function of citizens' moving costs and governments' factor prices; and
- interjurisdictional cooperation, determining the emergent scale of public goods as a function of organization costs.

This model is based on very weak assumptions about the knowledge of citizens and the benevolence of governments. Citizens are assumed to know nothing except their own personal experience living within a given jurisdiction – hence, dissatisfied citizens move out to random new jurisdiction. Local governments are assumed to be revenue-maximizing Leviathans, taking decisions about increasing the quality of public goods or of cooper-

ating with other jurisdictions solely on the basis of increasing their own rents (that is, difference between tax revenues and the cost of providing public services).

We have shown that, even under such harsh assumptions, jurisdictions will not increase taxes without increasing the quality of services, and they will respond to outflows of citizens by increasing the quality of services. Because moving costs are not zero, citizens will also respond to declines of quality, first by increasing their own participation in co-production activities and by means of voice, and only as a last resort by moving out.

We have also explored various inefficiencies that can emerge within the model. Shocks to factor prices can lead to downward spirals of quality and outflow migration. Both under-centralization and over-centralization vicious cycles are possible under certain conditions. Moreover, even without such vicious cycles, our Leviathan model does not necessarily lead to the level of centralization that the calculus of consent describes as the social contract optimum. This opens the door for further research into analyzing what types of constitutional constraints may force the Leviathan model to converge on the calculus of consent optimum.

NOTES

1. For example, in their seminal contribution to this literature, Klein and Leffler (1981), dissipate profits through non-price competition: firms invest in non-salvagable firm-specific assets, such as firm logos, expensive signs, and elaborate storefronts promoting the firm's name, that provide the greatest direct service value to consumers. These brand-name capital investments not only help alert consumers to the presence of the firm's product, but such investments also serve an important signaling role: they demonstrate the size of the sunk capital selling costs incurred by the firm and signal to the buyer the presence of price premia.
2. To solve for critical minimum of households, the manner in which jurisdictions form their expectations must be specified. Although we do not provide the details here, for a number of common-sense assumptions, we obtain the intuitive conclusion that the minimum number of households a jurisdiction must expect is a decreasing function of their probability of them moving out (see McPhail 2001 for proofs).
3. A more complex model of social capital would involve a nonlinear expression for H as a function of the individual h_is. The exact non-linear expression would describe the social structure of the jurisdiction, for example, if households i and j have a shared experience and relationship, the term $h_i h_j$ will appear in the expression of H. The simple linear form in the equation above does not include any social structure, that is, households act independently of one another.

REFERENCES

Alderman, L. and J. Kanter (2016), 'Europe's border checks become economic choke points', *New York Times*, 1 March.

Aligica, P.D. and P.J. Boettke (2009), *Challenging Institutional Analysis and Development: The Bloomington School*, London and New York: Routledge.

Aligica, P.D. and V. Tarko (2012), 'Polycentricity: from Polanyi to Ostrom, and beyond', *Governance*, **25** (2), 237–62.

Aligica, P.D. and V. Tarko (2013), 'Co-production, polycentricity, and value heterogeneity: the Ostroms' public choice institutionalism revisited', *American Political Science Review*, **107** (4), 726–41.

Allen, F. (1984), 'Reputation and product quality', *RAND Journal of Economics*, **15** (3), 311–27.

Bish, R.L. (1971), *The Public Economy of the Metropolitan Areas*, Wisbech: Markham.

Bish, R.L. and R.J Kirk (1974), *Economic Principles and Urban Problems*, Englewood Cliffs, NJ: Prentice-Hall.

Bish, R.L. and V. Ostrom (1979), *Understanding Urban Government: Metropolitan Reform Reconsidered*, Washington, DC: American Enterprise Institute for Public Policy Research.

Boettke, P.J. and P.T. Leeson (2012), 'Liberalism, socialism, and robust political economy', *Journal of Markets & Morality*, **7** (1), 99–111.

Boettke, P.J. and A. Marciano (2016), 'The distance between Buchanan's 'An Economic Theory of Clubs' and Tiebout's 'A Pure Theory of Local Public Expenditures'. New Insights based on an unpublished manuscript', GMU Working Paper in Economics No. 16-10, George Mason University, Fairfax, VA, accessed 2 March 2016 at http://papersssrncom/abstract=2729670.
Boettke, P.J., C.J. Coyne and P.T. Leeson (2011), 'Quasimarket failure', *Public Choice*, **149** (1–2), 209–24.
Boettke, P.J., J.S. Lemke and L. Palagashvili (2016), 'Re-evaluating community policing in a polycentric system', *Journal of Institutional Economics*, **12** (2), 305–25.
Boettke, P.J., L. Palagashvili and J. Lemke (2013), 'Riding in cars with boys: Elinor Ostrom's adventures with the police', *Journal of Institutional Economics*, **9** (4), 407–25.
Brandsen, T. and V. Pestoff (2006), 'Co-production, the third sector and the delivery of public services', *Public Management Review*, **8** (4), 493–501.
Brennan, G. and J.M. Buchanan (1977), 'Towards a tax constitution for Leviathan', *Journal of Public Economics*, **8** (3), 255–73.
Buchanan, J.M. (1987), 'Justification of the compound republic: the calculus in retrospect', *Cato Journal*, **7** (2), 305–12.
Buchanan, J.M. and G. Brennan (1980), *The Power to Tax*, Cambridge and New York: Cambridge University Press.
Buchanan, J.M. and G. Brennan (1985), *The Reason of Rules*, Cambridge and New York: Cambridge University Press.
Buchanan, J.M. and R.L. Faith (1987), 'Secession and the limits of taxation: toward a theory of internal exit', *American Economic Review*, **77** (5), 1023–31.
Buchanan, J.M. and C.J. Goetz (1972), 'Efficiency Limits of fiscal mobility: an assessment of the Tiebout Model', *Journal of Public Economics*, **1** (1), 25–43.
Buchanan, J.M. and G. Tullock (1962), *The Calculus of Consent*, reprinted 1999, Indianapolis, IN: Liberty Fund.
Caplan, B. (2001), 'Standing Tiebout on his head: tax capitalization and the monopoly power of local governments', *Public Choice*, **108** (1–2), 101–22.
Donahue, J.D. (1997), 'Tiebout? Or not Tiebout? The Market metaphor and America's devolution debate', *Journal of Economic Perspectives*, **11** (4), 73–82.
Engineer, M. (1990), 'Brennan and Buchanan's Leviathan models', *Social Science Journal*, **27** (4), 419–33.
Farrant, A. (2004), 'Robust institutions: the logic of Levy?', *Review of Austrian Economics*, **17** (4), 447–51.
Feller, W. (1968), *An Introduction to Probability Theory and Its Applications*, vol. 1, 3rd edn, New York: Wiley.
Geradin, D. and J.A. McCahery (2004), 'Regulatory co-opetition: transcending the regulatory competition debate', in J. Jordana and D. Levi-Faur (eds), *The Politics of Regulation*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 90–123.
Gintis, H. (1989), 'Power to switch: on the political economy of consumer sovereignty', in S. Bowles, R.C. Edwards and W.G. Shepherd (eds), *Unconventional Wisdom: Essays in Honor of John Kenneth Galbraith*, New York: Houghton-Mifflin, pp. 65–80.
Hirschman, A.O. (1970), *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*, Cambridge, MA: Harvard University Press.
Howell-Moroney, M. (2008), 'The Tiebout Hypothesis 50 years later: lessons and lingering challenges for metropolitan governance in the 21st century', *Public Administration Review*, **68** (1), 97–109.
Jones, A.H.M. (1986), *The Later Roman Empire, 284–602: A Social, Economic, and Administrative Survey*, 2 vols, Baltimore, MD: Johns Hopkins University Press.
Klein, B. and K.B. Leffler (1981), 'The role of market forces in assuring contractual performance', *Journal of Political Economy*, **89** (4), 615–41.
Kornai, J. (1986), 'The Soft budget constraint', *Kyklos*, **39** (1), 3–30.
Kossinets, G. and D.J. Watts (2009), 'Origins of homophily in an evolving social network', *American Journal of Sociology*, **115** (2), 405–50.
Kreps, D.M. (1990), *A Course in Microeconomic Theory*, Princeton, NJ: Princeton University Press.
Leeson, P.T. and J.R. Subrick (2006), 'Robust political economy', *Review of Austrian Economics*, **19** (2–3), 107–11.
Levy, D.M. (2002), 'Robust institutions', *Review of Austrian Economics*, **15** (2–3), 131–42.
Lowery, D. (1998), 'Consumer sovereignty and quasi-market failure', *Journal of Public Administration Research and Theory: J-PART*, **8** (2), 137–72.
Lowery, D. and W.E. Lyons (1989), 'The impact of jurisdictional boundaries: an individual-level test of the Tiebout model', *Journal of Politics*, **51** (1), 73–97.
Lyons, W.E. and D. Lowery (1989), 'Citizen responses to dissatisfaction in urban communities: a partial test of a general model', *Journal of Politics*, **51** (4), 841–68.
MacMullen, R. (1990), *Corruption and the Decline of Rome*, reprinted edn, New Haven, CT: Yale University Press.

McGinnis, M.D. (ed.) (1999), *Polycentricity and Local Public Economies: Readings from the Workshop in Political Theory and Policy Analysis*, Ann Arbor, MI: University of Michigan Press.
McGinnis, M.D. (ed.) (2016), 'Polycentric governance in theory and practice: dimensions of aspirations and practical limitations', paper presented at the Vincent and Elinor Ostrom Workshop in Political Theory and Policy Analysis, 14–17 December 2015, Indiana University, Bloomington, IN.
McGinnis, M.D. and E. Ostrom (2012), 'Reflections on Vincent Ostrom, public administration, and polycentricity', *Public Administration Review*, **72** (1), 15–25.
McGuire, M.C. and M. Olson (1996), 'The economics of autocracy and majority rule: the invisible hand and the use of force', *Journal of Economic Literature*, **34** (1), 72–96.
McPhail, E. (1997), 'Endogenous quality and intra-industry trade', working paper presented at the University of California, Riverside, CA, February 1998.
McPhail, E.A. (2001), 'Endogenous quality and intra-industry trade', PhD dissertation, University of Massachusetts, Amherst, MA.
Oakerson, R.J. and R.B. Parks (1988), 'Citizen voice and public entrepreneurship: the organizational dynamic of a complex metropolitan county', *Publius*, **18** (4), 91–112.
Olson, M. (1971), *The Logic of Collective Action: Public Goods and the Theory of Groups, Second Printing with New Preface and Appendix*, revised edn, Cambridge, MA: Harvard University Press.
Ostrom, E. (1976), *The Delivery of Urban Services: Outcomes of Change*, Beverly Hills, CA: Sage.
Ostrom, E. (1990), *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge: Cambridge University Press.
Ostrom, E. (1999a), 'Polycentricity, complexity, and the commons', *The Good Society*, **9** (2), 37–41.
Ostrom, E. (2005), *Understanding Institutional Diversity*, Princeton, NJ: Princeton University Press.
Ostrom, E. (2010), 'Beyond markets and states: polycentric governance of complex economic systems', *American Economic Review*, **100** (3), 641–72.
Ostrom, E., R.B. Parks and G.P. Whitaker (1978), *Patterns of Metropolitan Policing*, Pensacola, FL: Ballinger
Ostrom, V. (1991), *The Meaning of American Federalism*, San Francisco, CA: ICS Press.
Ostrom, V. (1999b), 'Polycentricity (part, 1 And, 2)', in M.D. McGinnis (ed.), *Polycentricity and Local Public Economies*, Ann Arbor, MI: University of Michigan Press, pp. 52–74 and 119–38.
Ostrom, V. and E. Ostrom (1977), 'Public goods and public choices: the emergence of public economies and industry structures', reprinted in V. Ostrom (1991), *The Meaning of American Federalism: Constituting a Self-Governing Society*, San Francisco, CA: ICS Press, pp. 163–97.
Ostrom, V., R.L. Bish and E. Ostrom (1988), *Local Government in the United States*, San Francisco, CA: ICS Press.
Ostrom, V., C.M. Tiebout and R. Warren (1961) 'The organization of government in metropolitan areas: a theoretical inquiry', *American Political Science Review*, **55** (4), 831–42.
Parks, R.B., P.C. Baker, L. Kiser, R. Oakerson, E. Ostrom, V. Ostrom et al (1981), 'Consumers as coproducers of public services: some economic and institutional considerations', *Policy Studies Journal*, **9** (7), 1001–11.
Pennington, M. 2011 *Robust Political Economy: Classical Liberalism and the Future of Public Policy* Cheltenham, UK ; Northampton, MA, USA: Edward Elgar Pub.
Rankin, J. (2016), 'EU's Schengen members urged to lift border checks to save passport-free zone', *Guardian*, 2 March.
Rogerson, W.P. (1983), 'Reputation and product quality', *Bell Journal of Economics*, **14** (2), 508–16.
Rogerson, W.P. (1987), 'The dissipation of profits by brand name investment and entry when price guarantees quality', *Journal of Political Economy*, **95** (4), 797–809.
Shapiro, C. (1982), 'Consumer Information, product quality, and seller reputation', *Bell Journal of Economics*, **13** (1), 20–35.
Shapiro, C. (1983), 'Premiums for High quality products as returns to reputations', *Quarterly Journal of Economics*, **98** (4), 659–79.
Tarko, V. (2015), 'Polycentric structure and informal norms: competition and coordination within the scientific community', *Innovation: The European Journal of Social Science Research*, **28** (1), 63–80.
Tiebout, C.M. (1956), 'A pure theory of local expenditures', *Journal of Political Economy*, **64** (5), 416–24.
Wagner, R.E. (2005), 'Self-governance, polycentrism, and federalism: recurring themes in Vincent Ostrom's scholarly oeuvre', *Journal of Economic Behavior & Organization*, **57** (2), 173–88.