

Against Gargantua: The study of local public economies

Chapter 1 from Vlad Tarko, *Elinor Ostrom: An Intellectual Biography* (Rowman Littlefield, 2017).



Elinor Ostrom (center), Roger Parks (left) and Diane Eubanks (right) analyze data on American law enforcement agencies (1977, Indiana University)

The presumption that economies of scale were prevalent was wrong; the presumption that you needed a single police department was wrong; and the presumption that individual departments wouldn't be smart enough to work out ways of coordinating is wrong. ... For patrolling, if you don't know the neighborhood, you can't spot the early signs of problems, and if you have five or six layers of supervision, the police chief doesn't know what's occurring on the street.

(Elinor Ostrom, interviewed by Zagorski 2006)

I have seen the ways that police officers serving an independent community, where local citizens have constituted it, deal with citizens. Citizens are treated differently when you live in a central city served by a metropolitan police department. Many of the officers in very big departments do not see themselves as responsible to citizens. They are on duty for specific hours and with an entirely different mentality ... When you are in a police car for eight hours with officers from a big department, you learn that they really do not know the area they are currently serving since they rotate so frequently. When I was in a police car with an officer from a moderately sized department, they would start telling me about the local community, where there are trouble spots, and where few problems occur. They watch trouble spots that they see potentially emerging. They would sometimes take a juvenile to their home in order to discuss problems they are observing. They do not put kids in jail the first time they observe behavior that is problematic. In the big cities, officers tend to charge juveniles who have been seen to commit small offenses right away. Many jails are overcrowded with juveniles in large cities. Problems of law enforcement in central urban districts have grown over time and are linked to the way urban governance has been shifted to ever-larger units.

(Elinor Ostrom, cited by Boettke, Palagashvili, and Lemke 2013)

The first big impact that Elinor Ostrom and her team have had was thanks to their empirical study of local public economies. To the extent that people follow politics, it is usually the national level politics that captures their attention. But, in many ways, the local level has a more direct impact upon our lives. The local level issues tend to gather our attention mainly when something goes terribly wrong: police brutality

in Ferguson, Missouri; poisoned drinking water in Flint, Michigan; the George Washington Bridge lane closure scandal in Fort Lee, New Jersey; the failure of public schools in Philadelphia; etc. In many ways we don't usually hear that much about local public economies, because, by comparison to national level policies, things actually tend to work much better. The Bloomington School's work on local public economies helps explain both why the local public sector usually works relatively well, and also why the failures that we observe have a specific pattern. When local politics fails, it does not fail at random.

The failure is usually a failure of scale: either due to an unfortunate top-down intrusion upon the local level or due to the inability of local communities to overcome the transaction costs involved in cooperating for producing larger-scale collective goods. In the relatively distant past, the second type of problem was much more prevalent than it is today. Historically, the emergence of the nation state acted as a means to solve the problems caused by over-decentralization. Nowadays, however, it is the top-down intrusion that usually causes most problems, which is why the Bloomington School, and this book, focuses more on it.

1.1 From UCLA to Indiana

The opening salvo of Bloomington School's foray into the study of local public economies was a theoretical paper written by Vincent Ostrom, Charles Tiebout, and Robert Warren, titled "The organization of government in metropolitan areas" (1961) (also included in McGinnis 1999b, chap. 1; and in V. Ostrom 1991b, chap. 6). This work was supported by the Bureau of Governmental Research at UCLA, the Haynes Foundation, and by the Water Resources Center of the University of California, but they were far from happy with the result. In fact, they had wanted the exact opposite conclusion. It was very common at the time to think that the consolidation of local administrations into larger centralized administrative units would increase efficiency, due to economies of scale and the elimination of overlapping efforts, as argued for instance by Anderson and Weidner's book *American City Government* (1950). They were hoping for a

study that would provide more academic support to these calls to centralize the administration in metropolitan areas. More specifically, they wanted something that supported the “Lakewood Plan”:

[Vincent] Ostrom was already interested by the implications of the 1954 incorporation of the city of Lakewood, California ... which, in an unprecedented political bargain, would contract out the majority of its municipal services with Los Angeles County. The “Lakewood Plan,” as it came to be called, quickly became the model for succeeding contract cities in Los Angeles and nationwide. (Singleton 2015)

But, instead of going along with the fashionable view at the time, Ostrom, Tiebout and Warren wrote a powerful defense of the benefits of decentralization and overlapping jurisdictions. Their intent was to use economic theory as a tool

for escaping political scientists’ “compulsion to want to superimpose a structure in such political situation[s].” As the research plan explained, in contrast with the political scientist: “The economist might apply a model of industrial organization and treat the interaction within the area as the operation of the market system, recognizing the existence of imperfect competition in an oligopolistic setting.” ... [Ostrom and Tiebout] sought to place the organization of metropolitan governments in an explicitly economic framework, where municipal services were bought and sold in a “quasimarket.” This theorizing, however, drew the ire of other political scientists in the bureau, resulting in Tiebout’s and Ostrom’s removal from the project and the eventual joint authorship of an article with Robert Warren instead. (Singleton 2015)

To make matters worse from the point of view of the consolidationists, the paper got published by the *American Political Science Review*, the top political science journal in the world, and became massively cited. Referring to the consolidationists as supporters of “Gargantua” probably didn’t help win them many favors either. The conflict with UCLA’s Bureau of Governmental Research eventually led both Vincent Ostrom and Charles Tiebout to leave the university. Their initial intention was to go beyond just theory, but the conflict with the Bureau prevented it:

The Lakewood Project—the study of the contract system in Los Angeles—will be approached in light of the model furnished by economics. Rather than as a limited examination of relations between political units, this study will evaluate the performance of the county government as a seller of goods and services, and of particular local units as the buyers of these services, as unions organized to meet the demand of consumers (the citizens). (“An Approach to Metropolitan

Areas,” Research Seminar Minutes, September 23, 1959, UCLA’s Bureau of Governmental Research, cited by Singleton 2015)

This empirical program will be revived in the 1970s by Elinor Ostrom and her students. On the theoretical side, “[Vincent] Ostrom wanted to adapt the Lakewood project research into a book, titled ‘A New Approach to the Study of Metropolitan Government’ … though it never materialized.” (Singleton 2015) He did, however, develop the polycentricity concept further (V. Ostrom 1972, 1991a). Robert Warren finished his dissertation under Vincent Ostrom, and published some of the Lakewood project findings together with his theoretical analysis (1966).

Historically, metropolitan areas emerged in an unplanned fashion as small towns gradually grew up to the point where they reached one another, and combined into one *de facto*, but not *de jure*, large urban area. The administrative organization of these metropolitan areas remained decentralized because of their history. The question was: Should they now be centralized under a larger hierarchical system? The *de facto* organization was “polycentric” reflecting the “multiplicity of political jurisdictions in a metropolitan area” (V. Ostrom, Tiebout, and Warren 1961). Should this be changed? Polycentricity became an important concept, and it was the topic of Elinor Ostrom’s Nobel Prize address (2010a), because it not only merely describes the *de facto* situation, but can also be used to understand why the system works relatively well, in the sense that it has desirable *emergent, unplanned properties*.

Polycentricity was first defined as a system of “many centers of decision making that are formally independent of each other” (V. Ostrom, Tiebout, and Warren 1961), but which, nonetheless, are forced to interact in both competitive and cooperative fashions, and may be embedded into larger systems: the decision centers “take each other into account in competitive undertakings or have recourse to central mechanisms to resolve conflicts”, and, as a result, “the various political jurisdictions in a metropolitan area may function in a coherent manner with consistent and predictable patterns of interacting behavior” (*Ibid.*).

Once they moved at Indiana University and started the Workshop, they were able to return to the initial idea. Initially, Elinor Ostrom was hired for teaching “Introduction to American Government” on Tuesday, Thursday, and Saturday at 7:30 a.m. “How could I say no?” she joked later (Zagorski 2006). But the position eventually changed to a tenured position, and when she started having PhD students in the 1970s, she told them to pick up any empirical topic except groundwater, as she was tired of that subject after having done her dissertation on it (this was only temporary). At the suggestion of Roger Parks, they chose to study police departments. This was not an entirely foreign subject to Elinor Ostrom. As a graduate student, she had been a part of Robert Warren’s team in Vincent’s Lakewood Project.

Building on Ostrom, Tiebout, and Warren’s theoretical paper and on the initial idea behind the Lakewood Project, a large number of empirical studies followed (Elinor Ostrom 1976b; Elinor Ostrom, Parks, and Whitaker 1978; Bish 1971; Bish and Kirk 1974; Bish and Ostrom 1979; V. Ostrom, Bish, and Ostrom 1988; McGinnis 1999b; see also Aligica and Boettke 2009 for more on the social philosophy behind these studies).

As Roger Parks recalls (included in McGinnis 1999b, 349):

In the spring of 1970, an intrepid band of graduate and undergraduate students, ably led by Lin Ostrom, set forth on the streets of Indianapolis and its close suburbs. Their task was to collect data relevant to the question of whether police services were better supplied by large, highly professionalized bureaus or by much smaller departments characteristic of most suburban United States. In the face of recent [then] and recurring recommendations for consolidation of police forces, other public services, and local governments in urban areas, this venture seemed quixotic, but it was instead quite productive. From it sprang a stream of Workshop research that kept many of our friends employed. This research, we believe, helped to change the tenor of debates over how to organize policing in urban areas and, more broadly, whether consolidation of local governments in metropolitan areas was the unmixed blessing perceived by its proponents.

Perhaps we get the clearest idea as to why these studies were so successful from Elinor Ostrom’s 1972 paper, “Metropolitan Reform: Propositions Derived from Two Traditions” (included in McGinnis 1999b, chap. 6).

This paper draws a series of empirically testable hypotheses about the efficiency of local public services from, on one hand, the consolidationist perspective and from, on the other hand, the “political economy”

perspective of V. Ostrom, Tiebout and Warren and the other public choice scholars (such as George Stigler¹, Mancur Olson², William Niskanen³ and Wallace Oates⁴), and then moves toward testing them empirically. The two perspectives are, hence, turned from being two grand conflicting *philosophical* visions to being simply two sets of hypotheses about how the world works, ready for scientific testing. Furthermore, as I discuss in more detail in chapter 2, the challenge of actually testing these hypotheses has shined a bright light upon the conceptual difficulties of trying to assess the quality of public services, and of evaluating what an “optimal” tradeoff between quality and cost might be (V. Ostrom and Ostrom 1977).

This approach, taking apart what may appear as high-level philosophical musings into their elementary components with the purpose of deriving empirically testable hypotheses, was one of the staples of Vincent Ostrom’s teaching. Filippo Sabetti, a graduate student at the Workshop in the late 1960s-early 1970s, who became a prominent first generation “Workshopper”, recalls that Vincent Ostrom’s graduate course “was based on readings derived largely from classical texts on America and the first generation of public choice scholars” (Sabetti 2011), covering books such as *The Federalist*, Tocqueville’s *Democracy in America*, Hobbes’ *Leviathan*, Buchanan and Tullock’s *Calculus of Consent*, and Olson’s *The Logic of Collective Action*. But Vincent Ostrom had a rather unusual approach to guiding students through these texts. His approach was “to pursue the logic of the propositions as testable hypotheses (at least by experience if not

¹ Stigler developed the theory of “regulatory capture” (see Bó 2006 for a review) – the idea that existing firms influence the regulatory process and obtain regulations that harm their competitors. This is what explains why existing firms often support government regulations of their industries.

² Olson (1965, 1982) developed and elaborated the consequences of the theory of “dispersed costs and concentrated benefits” – the idea that, because the costs of laws that grant special privileges to specific groups are spread out over a large number of people, each suffering only a minor harm, those harmed are not going to bother to organize to counteract the legislation; by contrast, those receiving the privilege do organize because each of them is individually receiving a large benefit.

³ Niskanen (1971) developed the public choice theory of bureaucracy, explaining why bureaucracies have unavoidable inefficiencies, no matter how well we may try to organize them.

⁴ Building on Tiebout (1956), Oates (1972, 1999) developed the theory of fiscal federalism, which tries to identify the efficient allocation of benefits and tax burdens at different levels.

by rigorous field research)" (Sabetti 2011).

We can see the success of the Bloomington School as emerging from this challenging approach to teaching that Vincent Ostrom employed, which was followed and expanded by Elinor Ostrom: "By early 1970s Lin's courses applying the Workshop's theoretical conceptions in rigorous fieldwork of police studies as well as in other modes of quantitative analysis and modeling, including game theory, had become part of the core curriculum, complementing Vincent's 'macro theoretic' approach with empirical studies and 'micro theory'" (Sabetti 2011). What made Vincent and Elinor Ostrom such successful intellectual and academic entrepreneurs, was not just that they had spotted neglected opportunities for research, but also that they developed their teaching curriculum and style to complement the research agenda.

It is perhaps worth pointing out that this approach to teaching stems from a particular philosophy of language, most prominently associated with John Searle and Michael Devitt (Searle 1969, 1999; Devitt and Sterelny 1999; see also Carnap 1950). Vincent Ostrom was particularly fond of Searle's (1969) *Speech Acts*. One of the key questions in the philosophy of language is "What is meaning?". To put it differently, when can we say that we truly understand something? Philosophers like Searle and Devitt argue that the meaning of a text (be it a single sentence or a larger text) is nothing but the set of empirical conditions that would have to hold in the real world for us to accept that what the text is saying is indeed true. In other words, to understand the meaning of a text, you have to be able to identify the ways in which the world would be different depending on whether the text were true or false.

As you can imagine, if this is your theory of language, you would indeed want to teach like Vincent Ostrom. As a professor, you would not believe that your students have truly understood a reading, be it as philosophical as it may be, unless they are able to derive empirically testable hypotheses from it. In turn, this has the great benefit of implicitly preparing the students to be empirical scientists who might be able

to tackle problems that many others would find too daunting. After all, once you've been forced to try to find testable hypotheses out of abstract philosophical texts, you may find it relatively easier to think empirically about more mundane scientific questions. Undoubtedly, Elinor Ostrom herself was one of the greatest students of this method, and what makes the Bloomington School rather unusual is that it successfully spans across the entire continuum from very high-minded social and political philosophy all the way to the most hard-nosed empiricism.

1.2 The complexity of public services

The provision of safety and protection is the most basic of governmental services. How should police departments be organized to provide the best services at the lowest cost? Elinor Ostrom coordinated a series of field studies of police services starting in Indianapolis, Chicago, Grand Rapids, Nashville-Davidson Country and St. Louis, followed by a much more extensive national level study covering 80 out of the 200 Standard Metropolitan Statistical Areas (SMSAs). This is still the largest study of police departments ever performed. As already mentioned, at the time, there were numerous calls to reform the organization of police departments by consolidating them. However, they discovered that “[n]one of the major national recommendations for change cite empirical evidence to support their contentions”, and, furthermore, they were based on large misconceptions about how police departments were actually organized (Elinor Ostrom, Parks, and Whitaker 1978, xxii). To make matters worse, the only areas that had been “frequently studied by national commissions, state and federal agencies, and social scientists” were the “[p]olice departments in the very large metropolitan areas”, despite “nearly half the nation’s people living outside the 16 largest metropolitan areas” (Elinor Ostrom, Parks, and Whitaker 1978, 7–8). Consequently, the recommendations for police “reform” were not just done *without* accurate information, but they were actually based on misleading and non-representative information gathered only from the largest metropolitan areas:

Figure 1: Cumberland County, NC



The New York City Police Department, the Chicago Police Department, the Los Angeles Police Department, and similar large city police agencies have been studied extensively. They are the models used explicitly or implicitly in judging other police departments across the country. But little has been written about the ways in which smaller police agencies work in relation to each other in small- and medium-sized urban areas across the country. (Elinor Ostrom, Parks, and Whitaker 1978, 2)

Furthermore,

Many observers have expressed concern about the existence of small departments. In their opinion these agencies cannot possibly provide the wide range of services offered by large central city police departments. Major changes in the law enforcement systems of metropolitan areas

have been proposed. These changes are being considered without much information (other than the number and size distribution of agencies) about how services are delivered. (Elinor Ostrom, Parks, and Whitaker 1978, 2)

In response to this challenge, Elinor Ostrom and her collaborators sought, first, to understand how these smaller police departments actually work, and, second, to evaluate how well they are doing their job. They analyzed three types of “direct services” – patrol, traffic control, and criminal investigation – and four types of “auxiliary services” – radio communication, adult pretrial detention, entry-level training, and crime labs. While the direct services impact citizens directly, the auxiliary services are factors of production used by the direct services. What they found was precisely what V. Ostrom, Tiebout, and Warren (1961) have intuited: such different services often have different optimal scales of production, and the organization of police departments involves complex delivery patterns, involving (1) autonomous local provision of some services, (2) cooperation between jurisdictions for larger scale problems, (3) alternate provision, based on predefined criteria, of the same service by more than one agency in the same jurisdiction, (4) only very rarely, duplication of the service provision by more than one agency without either cooperation or predefined criteria for alternation.

As an example of these complex patterns of delivery consider their in-depth analysis of the Cumberland county in North Carolina, the location of the Fayetteville metropolitan area (Figure 1). In this county, several jurisdictions are present: the military base at Fort Bragg, the Pope Air Force Base, Hope Mills, Spring Lake, Fayetteville, Fayetteville State University, and the rest the county.

With respect to patrol services, each of these areas has its own patrol, with the exception of the unincorporated part of the county, roughly covering everything east of I-95, which is patrolled by the North Carolina Highway Patrol.

The situation is much different with respect to homicides investigations. Criminal investigations in Fort

Bragg and in Pope Air Base “alternate” between the Federal Bureau of Investigations and the US Army Criminal Investigations Division or US Air Force Office of Special Investigations, depending on whether the homicides involve civilians or only military personnel (Elinor Ostrom, Parks, and Whitaker 1978, 42–44). In Hope Mills and Spring Lake the local police department cooperates with the Cumberland County Sheriff’s Department. In Fayetteville, including the university, the city’s police department has sole jurisdiction, while the rest of the county is serviced by the Sheriff.

The auxiliary services also showcase interesting examples of complex patterns of delivery (Elinor Ostrom, Parks, and Whitaker 1974, 1978, 50). For instance, the entry-level training is provided internally in the cases of the US Army Military Police, the US Air Force Security Police, the US Air Force Office of Special Investigations, the FBI, the North Carolina Highway Patrol, and the North Carolina State Bureau of Investigations. By contrast, the Fayetteville Technical Institute is providing entry-level training for the police departments in Hope Mills, Spring Lake, Fayetteville, Fayetteville State University, and Cumberland County Sheriff’s Department. Interestingly, the Spring Lake police department also gets some of its training from the Johnston Technical Institute – an example of duplication of services. But the competition between technical institutes is probably beneficial for keeping the quality of the training high.

In case of crime labs, while the Fort Bragg and the Pope Air Force Base have their own labs, the police departments in Fayetteville, Spring Lake, Hope Mills and the rest of the Cumberland county use either the Fayetteville-Cumberland County Crime Lab or the NC Bureau of Investigation lab, the two labs cooperating with each other (Elinor Ostrom, Parks, and Whitaker 1974 also included in McGinnis 1999b, chap. 11).

Similarly, in case of adult pretrial detention, the following arrangement exists between three producers:

The county sheriff provides adult detention facilities for all local police units. Each of the two military installations maintains its own detention facility. Military personnel taken into custody

by civilian departments are usually remanded to their base for detention. Civilians arrested on one of the military bases would be sent to the county jail for detention. Thus, the three producers can be viewed as alternative services restricted by clientele. (Elinor Ostrom, Parks, and Whitaker 1974)

This kind of analysis for one metropolitan area was replicated to all 80 metropolitan areas with similar results about complexity and cooperation across agencies. For example, although many police departments have their own crime labs, “[s]tate agencies supply lab services to police agencies in all but 1 of 80 SMSAs. Seven states (California, Connecticut, Kentucky, Massachusetts, Oklahoma, Ohio, and South Carolina) have two different state agencies that supply lab services to direct service police agencies” (Elinor Ostrom, Parks, and Whitaker 1978, 279). The bottom line is that, despite the conventional wisdom at the time that “having all services within the same department facilitates communication and coordination ... [i]n studying SMSAs with extensive division of services across agencies, we have, however, found considerable interdepartmental communication and coordination” (Elinor Ostrom, Parks, and Whitaker 1978, xxxv). In other words, one of the main arguments of consolidationists was shown to be based on a very poorly informed perspective about how metropolitan areas actually operate. The consolidationists held that

Fragmentation of police services is extreme ... Wasted energies are lost motion due to overlapping, duplication, and noncooperation are not the worst consequences of this fragmentation. Large areas of the United States – particularly rural communities and the small jurisdictions in or near metropolitan areas – lack anything resembling modern, professional police protection. (Committee for Economic Development, cited by Elinor Ostrom, Parks, and Whitaker 1974)

This was certainly *not* what Elinor Ostrom and her collaborators found to be the case. The consolidationists turned out to be mainly engaged in scare tactics and preconceived ideas with virtually no grounding in reality. The reality on the ground, as described above, was much better characterized as “polycentric”:

By “polycentric” I mean a system where citizens are able to organize not just one but multiple governing authorities, as well as private arrangements, at different scales. Each unit may exercise considerable independence to make and enforce rules within a circumscribed scope of authority

for a specified geographical area. In a polycentric system, some units are general-purpose governments, whereas others may be highly specialized [such as] ... special districts, private associations, or parts of local government. These are nested in several levels of general-purpose governments that also provide civil equity as well as criminal courts. (Elinor Ostrom, interviewed by Aligica 2003)

This type of polycentric organization is not restricted solely to police departments. It covers *all* aspects of public economies. As we'll see later in the chapter, we should expect, for a number of reasons, that *any* hierarchical arrangement of public economies will unavoidably be inefficient and slow to respond to changes, and the only way to have a responsive system is to have a polycentric system which addresses various issues at different scales. This means that, to the extent that our governments are democratically accountable, we should expect them to create a polycentric organization of public administration. To the extent that politicians and other "public entrepreneurs" face constraints in having to deliver good governance, they will be guided by the inherent logic of the complexity of collective goods to create polycentric arrangements. It is thus not an accident that we observe polycentric systems everywhere across the world where self-governance is allowed to operate as a guiding principle.

Ignoring polycentricity, partly because of misguided ideas about "reform", can potentially generate huge costs, especially due to the logic of bureaucracy which makes it exceedingly difficult to turn back from such "reforms" later, once they have failed. Elinor Ostrom begins her introduction to *The Delivery of Urban Services* by highlighting this very issue: "Failure, in many cases, leads to adoption of another program – one often based, as was the first, on inadequate analysis of the strategic behavior of the different actors. Failure seems to breed failure." (Elinor Ostrom 1976b, 7)

1.3 The rise and fall of community policing

What about the quality of police protection? As Elinor Ostrom (2010b) recalled, after comparing "police

departments serving similar neighborhoods within a metropolitan area ... [w]e never found a large department policing numerous neighborhoods that outperformed smaller departments within the same metropolitan area in regard to direct services to citizens". Furthermore, the "most efficient producers supply more output for given inputs in high multiplicity metropolitan areas than do the efficient producers in metropolitan areas with fewer producers" (E. Ostrom and Parks 1987, chapter 12 in McGinnis 1999b, 287).

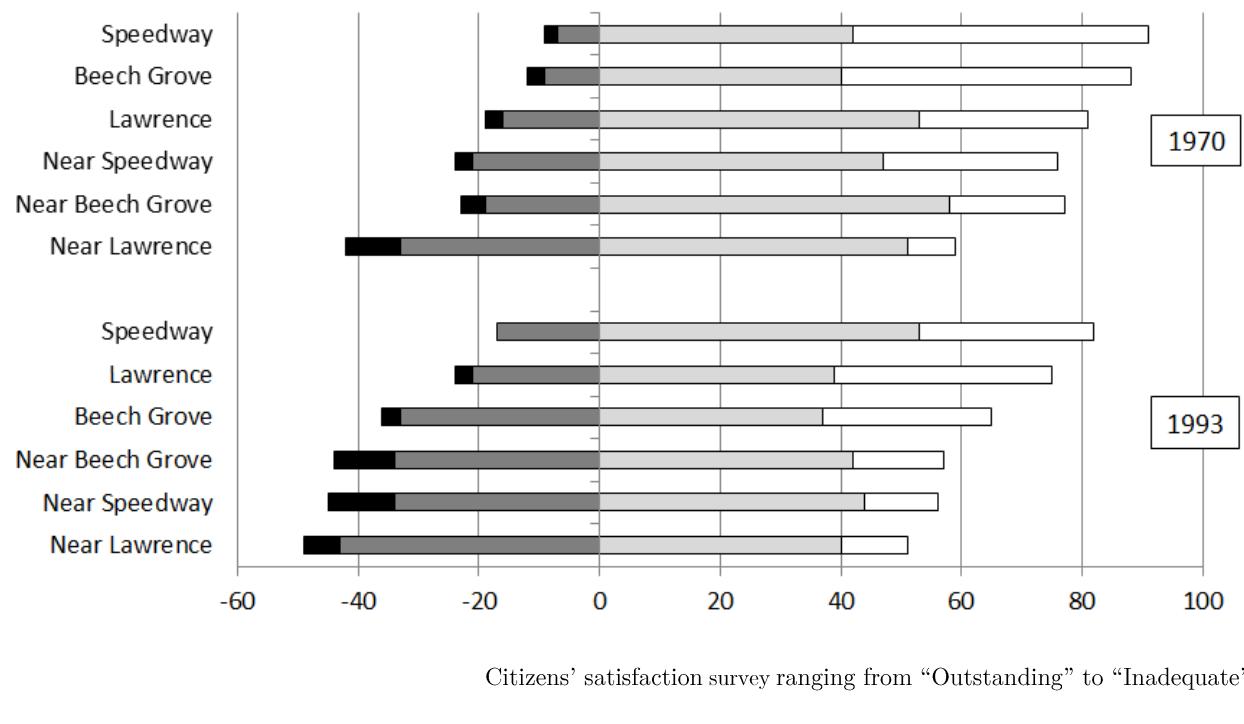
Table 1: Performance of consolidated vs. independent police departments

Variables	Independent communities	Indianapolis neighborhoods
<i>Experience variables</i>		
Victimization	+	-
Willingness to report victimization	+	-
Extent of police follow-up	+	-
Assistance	+	-
Promptness of response	+	-
Quality of assistance	=	=
Stopped as suspected offender	=	=
<i>Evaluation variables</i>		
Promptness	+	-
Crime trend	+	-
Potential for bribe taking	=	=
Police-citizens relations	+	-
General eval. of job being done	+	-

Legend: + is higher level of performance, - is lower level of performance, = is similar level of performance

(Source: E. Ostrom and Whitaker 1973, chapter 8 in McGinnis 1999b, 197)

Figure 2: Performance of independent vs. consolidated police departments



The task of measuring the performance of different police departments was a challenge in itself:

When conducting studies in the Indianapolis, Chicago, St. Louis, Rochester, and Tampa-St. Petersburg metropolitan areas, we solved the severe problem of measuring police performance by collecting performance data from interviews with a random sample of households served by small and large departments. Information was obtained about victimization, willingness to call the police, speed of police response, amount of police follow-up, satisfaction levels with police contacts, and general evaluations of the quality of policing in a neighborhood . . . (Elinor Ostrom 2000)

After comparing similar neighborhoods, the conclusion was entirely at odds with the received "wisdom" of consolidation:

The consistent finding from this series of studies is that small and medium-sized police departments perform more effectively than large police departments serving similar neighborhoods, and frequently at lower costs . . . Victimization rates tend to be lower, police

response tends to be faster, citizens tend to be more willing to call on police, citizens tend to more positively evaluate specific contacts with the police, and citizens tend to rate police higher across a series of other evaluative questions. Further, citizens living in small communities tend to be more informed about how to change local policies, tend to know more policemen serving their neighborhoods, and call the police more frequently to obtain general information than do citizens living in large cities. Citizens served by small departments tend to receive better services at lower costs than their neighbors living in the center city. Instead of being a “problem” for the metropolitan area, small departments frequently contribute to the improvement of police services in the area. (Elinor Ostrom 2000)

Table 1 and Figure 2 show a comparison of the satisfaction with police performance between six neighborhoods in the Indianapolis area (E. Ostrom and Whitaker 1973, Parks 1995, chapters 8 and 15 in McGinnis 1999b). Beech Grove, Lawrence and Speedway have independent small-scale police departments, while the nearby neighborhoods are part of the Indianapolis consolidated police. We see that across the board, the satisfaction with the police is greater in the smaller departments – indicative of the fact that smaller departments tend to be more responsive to citizens’ needs. We also see that, across the board, the satisfaction with the police departments has decreased from 1970 to 1993, but it has decreased more in the consolidated departments.

1.3.1 How reliable are citizens’ surveys?

At first, they received some pushback on their findings due to skepticism about using surveys:

Many scholars and public officials are uneasy about any reliance upon data collected from a survey of citizens about public agency performance. Citizens are thought by some to be uninformed and unable to give reliable perceptions and/or evaluations of service levels. Whether one agrees with this view or not (we obviously do not), reliance upon any single mode of measurement can lead to errors or biases in measuring performance or productivity. (Elinor Ostrom 1976a)

In a sense, the anti-self-governance bias is so strong that may experts would even deny that people know more about their local conditions than the remote experts.

To counteract this line of criticism, they designed other experiments comparing objective measures of roads

and street lighting with citizens' perceptions. Street lighting was somewhat straightforward to measure using a precision light-meter. With respect to roads, they have used a so-called "rough-o-meter" developed by the Urban Institute, which basically amounted to measuring how smooth it is to drive a car on a road. This was basically an early version of the "Street Bump" smart phone app that exists today, which uses people's phones to detect bumps in the road as they're driving, and which is now used by some city governments to collect information about potholes. They have also "developed an observation form and procedure which can be used by trained observers to record specific data about various aspects of street condition ... [including] among other items, measuring all potholes on a blockface with a 'yardstick pothole measurer'" (Elinor Ostrom 1976a). Elinor Ostrom recalled how

On Palm Sunday one year ago, I found myself walking down an Indianapolis street, carrying a yardstick and dashing out between passing cars to measure potholes in the street. Why would any sane person dash out onto a busy street to risk their life to measure some holes in the ground? I must confess that I asked myself that question several times that day and other days while I helped develop out "unobtrusive" measures of road conditions. (Elinor Ostrom 1976a)

The bottom-line is that the comparison between objective physical measures and subjective perceptions revealed a high degree of correlation, contrary to the critics who "assumed inaccuracy of citizen perceptions of service levels". Some differences existed between people, with "persons with more than a high school diploma, those over 45, those who have lived on a block more than five years, and those living on medium to short blocks tended to be more 'accurate' in their perceptions of road roughness" (Elinor Ostrom 1976a). Interestingly, the only perception flaw they could find was that citizens' misjudged the lighting conditions over the entire street to be similar to those in the close proximity to their own homes. The study basically established that citizens' perceptions were far more reliable than the critics were suggesting.

Turning again to the issue of evaluating police services, they noted that apparently more objective indicators are actually deeply flawed. Most police department evaluate themselves by looking at *inputs*, rather than

at outputs. “The internal records of most police departments consist mainly on workload data: rates of reported crime, traffic citations, and clearance of reported cases, for example. ... Consumers’ evaluations of the services they are receiving are not recorded at any point in routine police records.” (E. Ostrom and Whitaker 1973, chapter 8 in McGinnis 1999b, 180).

1.3.2 The failure of “community policing”

Partly as a result of these police and local public economies studies, a reform movement took shape in the 1980s. The calls for a move towards “community policing” became more popular. As a result of this

Police departments were advised to put police officers and community members in closer proximity by creating police “substations” and requiring police departments to have officers on foot patrol. These changes were intended to modify the existing conception of police as outsiders or threats and give better ground for community-police relationships. Another popular action was to hold community-police meetings so that the community could meet the officers and the officers could better understand the needs of the community. (Boettke, Lemke, and Palagashvili 2015)

At first glance, such reforms should have worked. But, instead, the results are generally mixed or negative, and the “reduction in crime [has been] primarily due to a trend towards larger police forces that has little to do with the adoption of any particular policing strategy” (Boettke, Lemke, and Palagashvili 2015).

These moves towards “community policing” have been, however, largely illusory or superficial. Part of the problem is that “community policing” has been adopted as a top-down policy and in many ways has been a form without content – more of a populist slogan than a genuine reform. A bigger part of the problem is that police activities have actually been centralized, with a greater impact of the federal level on local police departments, especially due to the War on Drugs and due to moves towards police militarization. In other words, whatever moves towards genuine community policing have been made, they have been swamped by further centralization:

The top-down approach to the establishment of community policing and the increasing reach of federal interventions into local law enforcement have prevented the emergence of true community policing as understood by Ostrom and her colleagues at The Workshop. Instead there has been a trend towards centralization and militarization of the police, shifting the focus away from the needs of the community and towards the homogenous goals of federal policy. (Boettke, Lemke, and Palagashvili 2012)

To make matters worse, the police-community meetings have gradually been bureaucratized, and as “the city began to send fewer beat police and more bureaucrats to community meetings ... participation waned as citizens began to feel that the purpose of the meetings had shifted from learning about citizen concerns to persuading the community to support traditional police action” (Boettke, Lemke, and Palagashvili 2015).

This failure of “community policing” was not so surprising to the Bloomington School itself. In fact Elinor Ostrom had warned precisely of this possible distortion of the idea right from the beginning. She noted that

The problems in obtaining an adequate knowledge about local situations have led several large-scale police departments to experiment with local commander systems and other arrangements to decentralize administrative control of neighborhood patrol forces. While this reform may increase direct supervision of patrolmen in the field and may lead to more effective coordination of their efforts within neighborhoods, *it may be expected to decrease the responsiveness to citizens of patrolmen serving these areas.* (E. Ostrom and Whitaker 1974, chapter 9 in McGinnis 1999b, 225, emphasis added)

In other words, community policing properly understood involves not just a decentralized mechanism for knowledge aggregation, but also – and this is critically important – allowing the *goals* of the local government to be determined by the local citizens. A decentralized police agency that takes orders from the federal government, and it is dependent on revenues and equipment on the central government, is not going to be responsive to citizens’ needs, and, hence, will not perform well in terms of most criteria of citizens’ satisfaction.

The idea that local police departments should be financially independent of central government, and, in order to be responsive to citizens, should be paid by the local community is known as “fiscal equivalence”

(Buchanan and Tullock 1962, 292–93; Olson 1969; V. Ostrom and Ostrom 1977). This was already familiar to Elinor Ostrom and collaborators, and they recognized its importance in connection to the problem of political representation:

Local control of the police would involve the establishment of formal structures of accountability to the public being served [rather than to the central government] as well as indirect internal supervision of patrolmen on the job. An effective means of establishing local control of the police in large cities might be to set up neighborhood districts to handle a variety of *locally confined public problems*. Such units would require some means of public selection of officials and *the authority to levy local taxes* and establish local ordinances. (E. Ostrom and Whitaker 1974, chapter 9 in McGinnis 1999b, 225, emphasis added)

The idea of fiscal equivalence is usually controversial because it implies that rich neighborhoods will have more resources than poor ones, and, hence better services. Elinor Ostrom's response to this is two-fold. On one hand, centralizing police departments and other local public services will not lead to improved outcomes for the poor neighborhoods – on the contrary, it further robs them of political representation. Centralized governments still respond more to richer constituencies, and, hence, absent local political representation, poorer neighborhoods will be even more likely to be left behind.

On the other hand, she notes that redistribution can be, to some extent, pursued as a distinct activity. We don't need to centralize the entire local public economy, with all its activities, just to distribute money to poorer neighborhoods. This being said, in order to avoid the Samaritan's Dilemma – the perverse distortion of the aid recipients' incentives away from self-reliance – she argued that redistribution should be kept limited to emergency situations: “Citywide forces could be utilized to supplement the needs of any local area *in times of emergency*. Redistribution to the poorer neighborhood districts within the large city could be provided from citywide as well as state and federal sources.” (E. Ostrom and Whitaker 1974, chapter 9 in McGinnis 1999b, 225, emphasis added)

It is also important to bear in mind that performance depends to a large extent on *how* the money is spent,

rather than on *how much* it is spent. For instance, they have found that “Chicago police spent 14 times the dollar amount that the independent police departments did – but despite this differential in expenditure, citizens in the smaller communities received the same or higher level of service” (Boettke, Palagashvili, and Lemke 2013).

This type of problem is most evident with respect to black neighborhoods. The study of police departments in Chicago found that “[b]lack citizens are among the constituents cited as least satisfied with the performance of local police and other public officials”. Redistribution did not help. On the contrary:

Redistribution of resources, itself, is not sufficient to bring about responsive police services. It appears that considerable resource redistribution is currently occurring within the city of Chicago. More resources are probably devoted to policing in black neighborhoods studied than are in derived revenue for such purposes from these areas. Residents of these neighborhoods, however, find police services no better and police somewhat less responsive than do village residents *despite* the much greater difference in resources devoted to policing. (E. Ostrom and Whitaker 1974, chapter 9 in McGinnis 1999b, 224)

They have found that because of so-called “preventive policing” there was, paradoxically, both too much policing and too little, as “police forces simultaneously increase the resentment of residents and diverts manpower from other activities such as answering calls and investigating the many crimes that do occur in the ghetto” (Ibid., 207). Interestingly, the negative attitudes of the black citizens towards the police were inversely related to their income: “black respondents of higher income levels tended to be less likely to give high ratings to police than black respondents of lower income levels” (Ibid. 206). The bottom line is that in Chicago, as well as in other cities, “[p]olice seems to be failing to serve residents of many black neighborhoods in US cities” (Ibid., 207), and the main reason is the subversion of genuinely local public economies by top-down control.

Since the Bloomington School did these studies, the situation seems to have deteriorated. For example, in Chicago today, a recent study by the Chicago Police Accountability Task Force, with members appointed

by Mayor Rahm Emanuel, has revealed a shocking situation (Lopez 2016). The report found that black residents were far more likely to be “[s]topped without justification, verbally and physically abused, and in some instances arrested, and then detained without counsel”. Furthermore, out of “404 police shootings between 2008 and 2015 … [a]mong the victims, 74 percent were black, even though black people make up just 33 percent of Chicago's population. … Of the 1,886 Taser uses between 2012 and 2015, 76 percent of those hit by stun guns were black.” This decades-long history of police racism significantly worsened due to the War on Drugs and police militarization (Boettke, Lemke, and Palagashvili 2012, 2015; Coyne and Hall-Blanco 2016), and it is now even more difficult to change: “False arrests, coerced confessions and wrongful convictions are also a part of this history. Lives lost and countless more damaged. These events and others mark a long, sad history of death, false imprisonment, physical and verbal abuse and general discontent about police actions in neighborhoods of color.” (Lopez 2016)

Back in 1974, Elinor Ostrom and Gordon Whitaker analyzed a variety of reform proposals, including calls for the increased “professionalization of the police force” and access to larger funding. Their conclusion was that “[p]olice effectiveness depends, in part, on police understanding the nature of the community being served and police openness to suggestions, criticism, and complaints” (E. Ostrom and Whitaker 1974, chapter 9 in McGinnis 1999b, 224). As a result, “[c]ommunity control of police may, thus, provide an institutional framework for the effective expression of black citizen demands for impartial police service. … Professionalism alone does not appear to provide sufficient controls so that police will be responsive to their needs for protection and respect. Community control places that responsibility on the people themselves and provides them with the mechanisms by which to exercise it.” (Ibid., 225-6). Unfortunately, this advice was not only not taken, but, instead, the phrase “community policing” was distorted beyond recognition to mean almost the exact opposite to what the Bloomington School recommended. As Vincent Ostrom (1991b, 1997) would increasingly complain over the course of his life, the Orwellian distortion of language is a

significant, under-studied phenomenon in public choice.

1.4 The impossibility of efficient hierarchical public economies

The most important point made by V. Ostrom, Tiebout and Warren (1961) is that, by analyzing the nature of public services that a metropolitan public administration has to solve, we are led to the conclusion that a consolidated, hierarchical administration would *unavoidably* lead to massive inefficiencies because *the administrative units operate at rigid scales, while the scale of public issues are varied and always changing*.

The idea behind this impossibility theorem is fairly simple: All complex societies face numerous collective issues, which, by their nature, occur at various scales, and, to make matters more complicated, these scales change all the time as a result of technology and other social processes. By contrast, the public administration units are relatively few in number (compared to the overall number of issues) and have relatively rigid geographical scales. As such, a given administrative unit is always faced with challenges that do not properly fit its administrative scale. This is true for administrative units at all scales. In a hierarchical system, when the scale of the problem is larger than the scale of administrative unit A, the responsibility for solving the problem goes to the larger administrative unit B, at a higher level. But B's scale can never fit exactly *all* the problems that are larger than A's scale. As such, the hierarchical system is bound to be rife with inefficiencies because A has only *one* higher-level unit, B. To make matters worse, the logic of bureaucracy leads all administrative units to expand beyond their proper scope (Niskanen 1971).

By contrast, polycentric governance assumes that small scale administrative units can organize on a quasi-ad hoc basis to address *some* of the larger scale problems (but not others). Different larger scale problems are addressed by different configurations of smaller scale units. We have seen this earlier. In their study of police departments across 80 metropolitan areas in the United States, Elinor Ostrom and her collaborators

have found that police departments cooperated in precisely this diverse manner to take advantage of economies of scale with respect to criminal investigations, adult pretrial detention, and auxiliary services (such as crime labs and entry-level training) (E. Ostrom, Parks, and Whitaker 1978). As Ostrom, Tiebout and Warren (1961) put it:

[T]he statement that a government is “too large (or too small) to deal with a problem” often overlooks the possibility that the scale of the public and the political community need not coincide with that of the formal boundaries of a public organization. Informal arrangements between public organizations may create a political community large enough to deal with any particular public’s problem. Similarly, a public organization may be able to constitute political communities within its boundaries do deal with problems that affect only a subset of the population. It would be a mistake to conclude that public organizations are of an inappropriate size until the informal mechanisms, which might permit larger or smaller political communities, are investigated.

These “informal mechanisms” are the alternative to the rigid hierarchical organization. We can also reframe the Ostrom-Tiebout-Warren impossibility theorem as an unavoidability theorem: Everywhere where politics works fairly well, i.e. where the public sector is fairly responsive to citizens’ needs and desires, we are bound to observe polycentric governance, rather than hierarchical governance. Vincent Ostrom stressed that the whole range of human affairs is polycentric (V. Ostrom 1991, 224). Markets, science, common law, competitive governance are standard examples of complex polycentric systems. But we also need to emphasize that this is just the beginning of the analysis. Not all polycentric systems are equally efficient. As we shall see in more detail in chapter 2, Elinor Ostrom has shown that not all societies successfully self-govern, but only those that happen to zero-in on a number of so-called “design principles”.

Let us look more carefully at the reasons why, unlike polycentric governance, hierarchical public administration would be bound to be rife with inefficiencies.

1.4.1 Control over the cause of the problem

The cause of the problem must be under the control of the administrative unit. For example, a water pollution problem caused by a factory up the river, outside the jurisdiction of the local authority, cannot be solved by that local authority. This is often a real problem. In the 1960s, when the Vincent Ostrom and his colleagues were writing, the following was happening:

Pasadena, for example, is subject to severe smog attacks, but the city's boundary conditions do not cover an area sufficient to assure effective control of the appropriate meteorological and social space that would include the essential variables constituting a "smogisphere" of southern California. None of the separate cities of southern California, in fact, can encompass the problem. Instead, county air pollution control districts were organized for the Los Angeles metropolitan community. The failure of even those counties to meet adequately the criterion of effective control has led the Californian state government to assume an increasingly important role in smog control. (V. Ostrom, Tiebout, and Warren 1961)

The same holds for positive externalities. If an administrative unit is providing a public good it has to have a way of preventing free-riding. For instance, if one group acts responsibly in order to avoid overfishing, the problem is solved only if other groups are also prevented from overfishing. Mirroring what Buchanan and Tullock (1962, chaps. 5–6) would famously write a year later in the *Calculus of Consent*, they summarize: "A function of government, then, is to internalize the externalities – positive and negative – for those goods which the producers and consumers are unable or unwilling to internalize for themselves" (V. Ostrom, Tiebout, and Warren 1961).

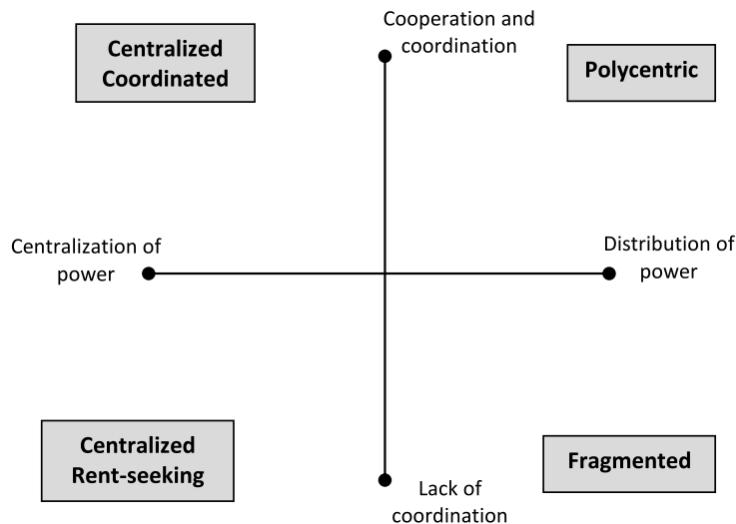
This idea is far more general than just the concern with metropolitan governance. For example, in her studies of the common-pool resources management problems, Elinor Ostrom encountered again and again the importance of polycentricity, and she is

urging readers to think more positively about complex, polycentric systems of governance that are created by individuals who have considerable autonomy to engage in self-governance. *Given the wide variety of ecological problems that individuals face at different scales, an important*

design principle is getting the boundaries of any one system roughly fit the ecological boundaries of the problem it is designed to address. Since most ecological problems are nested from very small local ecologies to those of global proportions, following this principle requires substantial investment in governance systems at multiple levels – each with some autonomy but each exposed to information, sanctioning, and actions from below and above. (Elinor Ostrom 2005a, 258)

Notice the similarity to the challenges to metropolitan governance. And as there, we encounter the same intellectual difficulty here: “one of the important threats is the effort to impose uniform rules and large boundaries on systems so they are more comprehensible to academics and policymakers” (Elinor Ostrom 2005a, 257).

Figure 3: The distinction between polycentricity and fragmentation



(Source: Pahl-Wostl and Knieper 2014)

Polycentricity is quite different from simply decentralization. Large scale problems require large scale solutions. The point isn't to *reduce* the scale of all administrative units, but to *fit* the scale of the administrative unit to the scale of the issue. This being said, the confusion between polycentricity and decentralization is somewhat understandable considering the fact that, in our world, over-centralization is far more common than under-centralization. As such, in practice, the concern for fitting the administrative

unit to the scale of the problem often amounts to calling for decentralization – as we have seen in the example of policing.

One simple way to understand the difference between polycentricity and mere decentralization (and thus fragmentation) is to think in terms of the distinction proposed by Pahl-Wostl and Knieper (2014) and illustrated in Figure 3.

1.4.2 Accurately measuring the demand for public goods and the opportunity cost of providing them

When the government is trying to fix a problem, the cost of the problem must be properly evaluated, for otherwise the administration might spend either too many too few resources for tackling it. But public goods are notoriously difficult to evaluate because everyone is expecting for others do shoulder the costs. For example, surveys are a somewhat imperfect method of assessing cost because people might complain about everything, far beyond the resources they are willing to contribute to address the problems. To put it differently, when somebody else is supposed to fix the problem, it's their job to find the resources to do it, our job is just to complain.

There is a more fundamental reason why the true demand for public goods is inherently difficult to evaluate:

A decision to buy a particular good or service reflects willingness to forgo all other opportunities for which the money could have been used. An expression of demand in a market system always includes reference to what is forgone as well as what is purchased.

The articulation of preferences in the public sector often fails to take account of forgone opportunities. ... Because most public goods and services are financed through a process of taxation involving no choice, optimal levels of expenditure are difficult to establish. (V. Ostrom and Ostrom 1977, 185)

V. Ostrom, Tiebout, and Warren (1961) refer to this issue as the problem of “packageability”, and note

that “private goods, because they are easily packageable, are readily subject to measurement and quantification. Public goods, by contrast, are generally not so measurable.” What they mean by this is that with private goods each person gets their own separate, individual good, while with public goods everyone gets the same shared good. This may make coercion necessary, in order to prevent free-riding, but

[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. (V. Ostrom and Ostrom 1977, 175)

What this means is that “alternative mechanisms to prices are needed for articulating and aggregating demands into collective choices that reflect individuals’ preferences for a particular quantity and/or quality of public goods or services” (V. Ostrom and Ostrom 1977, 175). But all such alternative mechanisms suffer from very serious problems. E. Ostrom and Whitaker (1973, chapter 8 in McGinnis 1999b, 179) noted this problem with respect to rigorously evaluating the outputs of police departments: “Police departments characteristically provide all services without consumer charges, even though some similar services are provided privately. Thus, the market value of police output cannot be obtained.”

To make matters worse, a standard result in public choice is that a person’s decisions as part of a collective, e.g. when they vote, are far less careful than their personal decisions, because one has less control over the outcome (Buchanan 1954; Olson 1965, 1982, chap. 2; Caplan 2008). For instance, regardless of the candidate with whom you personally vote, the same candidate will win the election, so there’s no point to regret your decision too much afterwards. By contrast, if one buys the “wrong” car, one doesn’t just have to live with it, one can also justifiably regret not having made a different decision. This leads to troubling conclusions with respect to how much attention people pay to costs of publicly provided goods.

1.4.3 Fiscal equivalence and redistribution

Mirroring Buchanan and Wagner's concerns in *Democracy in Deficit* (1977), Vincent and Elinor Ostrom note that

Costs must be proportioned to benefits if people are to have any sense of economic reality. Otherwise beneficiaries may assume that public goods are free goods, that money in the public treasury is “the government’s money”, and that no opportunities are forgone in spending that money. When this happens, the foundations of a democratic society are threatened. (V. Ostrom and Ostrom 1977, 186–87)

This problem of populist politicians promising “free goods” is of course perpetually occurring, and the voting public is more or less willing to let itself swayed by this illusion. Vincent and Elinor Ostrom’s note that an imperfect solution to this potentially large problem is the idea of “fiscal equivalence” (Olson 1969). In the *Calculus of Consent* chapter on the problem of pressure groups, Buchanan and Tullock describe this possible solution as “requir[ing] that individuals and groups securing differential benefits also bear the differential costs” (Buchanan and Tullock 1962, 291). As we shall see in chapter 4, this concept of fiscal equivalence is one of the major “design principles” for a sustainable social system, namely the “Fairness principle: Proportionality between the benefits and costs of various actors”.

A problem with fiscal equivalence, as both Buchanan & Tullock and Vincent & Elinor Ostrom recognized, is that it cannot deal very well with “cases where over-all redistribution cannot be put aside” (Buchanan and Tullock 1962, 292). Consider the following example:

[S]uppose that the issue confronted is that of providing federal funds to aid the depressed coal-mining area of West Virginia. For such a measure the levy of special taxes on citizens of West Virginia would be self-defeating. Nevertheless, it is relatively easy to see that, if such aid is to be financed out of general tax revenue, a veritable Pandora’s box may be opened. Depressed fishing villages along the Gulf coast, depressed textiles towns in New England, depressed automobile production centers in Michigan, depressed zinc-mining areas in Colorado, etc., may all demand and receive federal assistance. As a result excessive costs will be imposed on the whole population. (Buchanan and Tullock 1962, 292–93)

The question is thus, how can we organize the system such that “[g]enuinely depressed areas, considered as such by the whole population, would tend to be provided with assistance without at the same time opening up the whole set of grants to areas not considered to be deserving of assistance” (*Ibid.*)? We can think of two possible solutions to this problem.

The first solution is to draw a lesson from how insurance companies deal with moral hazard. For example, in order to avoid people indulging in too many unnecessary and expensive medical procedures, insurance companies require copayment. By requiring patients to pay a fraction of the cost themselves, the insurance companies try to introduce considerations of cost in the patients’ decision as to whether to have a particular procedure. The same type of approach can be used to address the above problem, with the federal level requiring a certain level of copayment from the local level.

For example, this is how the European Union’s structural funds work. In order to access these funds, usually provided to poorer members of the union by the richer members for specific projects, the recipient is required to co-sponsor the project. You can understand this as implicitly forcing the recipients to rank the importance of various possible projects, and ask for structural funds help only for the most important. Some of the more effective Eastern European governments, such as Poland, have managed to organize in this fashion relatively well, which has led to a high “absorption rate” of these funds (of the total funds available in principle, recipient countries usually manage to receive only a fraction, because they cannot organize to provide the copayment). They have used such help to fund relatively important goods, such as building roads. Other, less effective Eastern European governments, such as Romania, have been far less successful in absorbing the structural funds, which can be understood as a failure to organize collectively to rank the importance of various possible projects and co-sponsor only the most important ones.

On the same grounds, Elinor Ostrom was critical of interventions in developing countries which did not

take into account how the intervention ends up distorting the recipients incentives in perverse ways, what is known as the Samaritan's Dilemma (Buchanan 1975; Gibson et al. 2005):

Showering a region with funds is a poor investment if that serves mainly to bolster political careers and builds little at the ground level. It makes more sense to invest modest levels of donor funds in local projects in which the recipients are willing to invest some of their own resources. If the level of external funding becomes very large without being strongly tied to a responsibility for repayment over time, local efforts at participation may be directed more at rent seeking than at productive investment activities. (Elinor Ostrom, interviewed by Aligica 2003)

Another possible solution stems from the analysis of the problem in terms of "concentrated benefits and dispersed costs". Buchanan and Tullock's "Pandora's box" is opened because each special interest gets its concentrated benefit, while everyone gets stuck with the dispersed cost. For each case, the dispersed cost upon one individual is very small, but it adds up. Paradoxically, everyone can end up worst off because the amount they end up paying in total, a little bit by little bit, is lower than what they receive. Buchanan and Tullock thought that one possible institutional solution to this problem would be to organize the aid relationship on an equal size criterion: in order to assure that the recipients and the providers have equal bargaining power, they should be roughly the same in size. "For example, if the designed aid to West Virginia were to be collected from special taxes levied on Oklahoma only, then we could be assured that roughly balancing political forces would determine the final outcome" (Buchanan and Tullock 1962, 293). This sounds unusual, and, indeed, it does not reflect the hierarchical approach currently in use. It would describe instead a polycentric mutual aid system, in which providing assistance is paired with effective monitoring and assessment of genuine need. The civil society mutual aid societies that existed prior to the rise of the welfare state had precisely such a polycentric organization, but this civil society system was largely replaced by the rise of the welfare state (Beito 2000). To some extent, such civil society organization still exist in the realm of faith-based initiatives (McGinnis 2008, 2010).

1.4.4 The separation of production and provision

One of the recurring themes in the Bloomington School is the fact that public services and goods that are “consumed collectively” can be produced by a variety of methods. Unlike the simplified perspective in most accounts of public economics, according to which public services are assumed almost by definition to be produced by a government agency, the Ostroms and their collaborators documented many hybrid institutions that do not fit well in either purely “state” or purely “market” categories. In contrast to this conflation between the public sphere and government,

the work done at the Workshop demonstrates that public services need not be provided by a central government or the state. Many streets, roads, and other thoroughfares, fire protection, police services, and other such services may be arranged by local communities. These arrangements may rely on private entrepreneurs, but under terms and conditions that are communally specified. ...

We need not think of “government” or “governance” as something provided by states alone. Families, voluntary associations, villages, and other forms of human association all involve some form of self-government. Rather than looking only to states, we need to give much more attention to building the kinds of basic institutional structures that enable people to find ways of relating constructively to one another and of resolving problems in their daily lives. (Vincent Ostrom, interviewed by Aligica 2003)

One way to make sense of hybrid public-private arrangements is to think about the distinction between *provision* (paying for the good) and *production* (creating the good). As noted by Oakerson and Parks (2011), “[o]ne key insight of V. Ostrom, Tiebout, and Warren (1961) almost forty years ago was that public provision did not require public production by the same governmental unit. Indeed, all governments provide services to their citizenry that they do not produce in-house.” Oakerson and Parks (2011) also develop a more complex analysis of “provision”. They note that, beyond just the decision to subsidize the production of a particular good, provision may also include public decisions about “what quantities of each service to provide and what quality standards to apply, and how to arrange for and monitor production” (*Ibid.*). Such

decisions about whether “to contract out and what to produce in-house is a city-specific decision”, which “requires careful attention to the nature of specific public goods and services and the local market for their procurement” (*Ibid.*).

The institutional complexity of hybrid systems can be even greater. Elinor and Vincent Ostrom (1977) note that a government may obtain the desired public goods by a variety of methods such as:

Operating its own production unit. E.g. a city with its own fire or police department.

Contracting with a private firm. E.g. a city that contracts with a private firm for snow removal, street repair, or traffic light maintenance.

Establishing standards of service and leaving it up to each consumer to select a private vendor and to purchase service. E.g. a city that licenses taxis to provide service, refuse collection firms to remove trash.

Issuing vouchers to families and permitting them to purchase service from any authorized supplier. E.g. a jurisdiction that issues food stamps, rent vouchers, or education vouchers, or operates a Medicaid program.

Contracting with another government unit. E.g. A city that purchases tax assessment and collection services from a county government unit, sewage treatment from a special sanitary district, and special vocational education services from a school board in an adjacent city.

Producing some services with its own unit, and purchasing other services from other jurisdictions and from private firms. E.g. a city with its own police patrol force, which purchases laboratory services from the county sheriff, joins with several adjacent communities to pay for a joint dispatching service, and pays a private ambulance firm to provide emergency medical transportation.

These kinds of arrangements, which are all quite common, allow a much more flexible public sector, but they do not fit very well in simple economic categories assuming clear distinctions between private and public, markets and governments. They lead to what Richard Wagner (2014) has called “entangled political economy”.