# Polycentricity and the Meaning of Self-Governance: The Role of the Calculus of Consent in Elinor Ostrom's "Design Principles"

Vlad Tarko

George Mason University, Economics Department, Mercatus Center

#### Abstract

Why are certain issues addressed by markets, others by politics, and others by courts? Why do hybrid institutions such as independent regulatory agencies, judicial review, and private arbitration courts also exist? Why are certain issues addressed at the local levels, while others at the federal level? This paper explains how to use the calculus of consent model to answer such questions, and argues that the calculus of consent provides the general framework for explaining all institutional diversity. Institutional efficiency is defined as the difference between the de facto decision rule, as it exists thanks to actual institutions, and the optimal calculus of consent rule. To analyze institutional change and political reform, one looks at whether various available collective choice mechanisms can in fact change the de facto decision rule, bringing it closer to the optimal rule.

*Keywords:* calculus of consent; polycentricity; self-governance; majority rule *JEL Classification:* D02, D71, D74, D85, H11

Prepared for delivery at the Workshop on the Ostrom Workshop (WOW5) conference, Indiana University Bloomington, June 18-21, 2014.

(c) Copyright 2014 by the author

# 1 Introduction

Markets, democratic politics, and courts can be seen as varieties of non-violent methods for managing or preventing conflicts, and for distributing resources peacefully. But why are certain issues addressed by markets, others by politics, and others by courts? Why do hybrid institutions such as independent regulatory agencies, judicial review, and private arbitration courts also exist? Why are certain issues addressed at the local levels, while others at the federal level? The economic literature on conflict emphasizes *why* conflict resolution institutions tend to be created, and illustrates how some specific institutions are accomplishing this task (Tullock, 1974; Nye, 1997; Hirshleifer, 2001; Fearon, 1995; Filson and Werner, 2002; Reiter, 2003; Smith and Stam, 2004; Powell and Stringham, 2009; Leeson, 2006; Leeson et al., 2006; Arena, 2010). What is usually left out is the question about institutional diversity. What underlining factors explain not just the emergence of institutions limiting conflict and securing a certain level of prosperity, but also the observed diversity of those methods for conflict prevention and resolution?

To some extent, such questions have been addressed by authors like Fuller (1978), Ostrom (1987), Ostrom (2005) or Aligica (2014), but no general theory of institutional diversity has been proposed so far. Djankov et al. (2003) come closest to providing such a framework. They introduce a method of institutional analysis based on what they call the "institutional possibilities frontier" (IPF), which sets a trade-off between the costs due of private disorder and the costs of authoritarian rule. In its original form, the IPF construct is useful for dealing with a special case of the problem, the problem of "expert rule", but cannot deal well with the problem of institutional diversity in general. Moreover, at least so far, their approach is to a large extent only a device for creating "just so" stories. For it to be more than that we need a method for determining the IPF.

The main theoretical claim of this paper is that the calculus of consent model (Buchanan and Tullock, 1999; Wagner, 1988), can be used to get a better and more precise understanding of polycentricity, and, as a result, it provides the broad conceptual key for understanding institutional diversity — the key to explaining *all* institutions and not just the American democracy. In particular, it can be used to obtain a less ad hoc perspective on the IPF, while at the same time generalizing the idea to apply to more than just the problem of expert rule.

One can understand the role of the calculus of consent in institutional analysis, and its generality, by thinking of it as nothing but Coase's theorem with positive transaction costs, along the lines of Coase's (1960) own approach in the original paper and of the Wicksellian approach to political economy (Buchanan, 1987; Wagner, 1988; Besley, 2007; Boettke, 2012, chapter 17). The zero-transaction costs Coase theorem simply describes the manner in which all market failures would be internalized by bargaining in an idealized environment in which time constraints are not an issue. Given enough time, bargaining would indeed be sufficient to create the most efficient institutions. Because conflict is wasteful, people have an incentive to find ways to eliminate it (Tullock, 1974; Hirshleifer, 2001; Arena, 2010). Adding the positive transaction costs of bargaining to the Coasian model, i.e. allowing for realistic time constraints, amounts to acknowledging the existence of collective decision costs. But Buchanan & Tullock take a step further from Coase by assuming that, precisely because *one-on-one bargaining* would be very costly, communities create *other collective choice institutions* that approximate the hypothetical outcome of bargaining. This is the meaning of Buchanan & Tullock's idea of "conceptual unanimity".

In "The Quest for Meaning in Public Choice", Ostrom and Ostrom (2004) pointed out that their research program got off the ground precisely thanks to *The Calculus of Consent* which "gave us basic tools for acquiring some analytical leverage in addressing particular problems that people are required to address about public affairs". Indeed, the work of the Bloomington School can be understood as nothing but a detailed exploration into the mechanisms by which people devise institutions for lowering decision costs and/or external costs, i.e. for increasing the efficiency of their institutions thanks to the work of "public entrepreneurs" (Oakerson & Parks 1988). In Ostrom & Ostrom's (2004) words, Buchanan & Tullock's "principle of conceptual unanimity gave meaning to what [we] had observed and what was accomplished". For example, when Elinor Ostrom studied water management in West Basin, California, she found that adjudicating water rights was driven by intuitive considerations of "equity jurisprudence", which can be understood as seeking "to achieve conceptual unanimity in establishing the nature of the problem, in adjudicating water rights, in formulating the rules that were constitutive of water user associations, the way they related to one another, and in monitoring performance". Numerous other studies went along very similar lines, and with similar conclusions (Ostrom, 1990, 2005; Ostrom et al., 1994; McGinnis, 1999a,b).

We can interpret the idea of effective collective-choice arrangements to mean that the actual decision rules prevalent in the community for various issues are close to the calculus of consent optimal rules for those issues. Successful institutions are those that approximate the calculus of consent optimal decision rule. As Vincent Ostrom (1997) has put it, although exploitation and power relations persist, "I have come to the conclusion, however, that democratic societies are necessarily placed at risk when people conceive of their relationships as being grounded on principles of command and control rather than on principles of self-responsibility in self-governing communities of relationships" (p. 4). Note that he is making primarily a *positive* claim about what kinds of institutional arrangements tend to survive over the long-term. Although, arguably, Stigler (1992) is going too far in assuming that all "enduring institutions" are efficient in all their details, one can still make an evolutionary argument about a broad and imperfect tendency of societies to zero-in on certain efficient "design principles" (Wilson et al., 2013). The Calculus of Consent has thus played a critical role in the origins of the Bloomington School providing one of the essential conceptual guidelines for their research program. However, we hold that the fruitful mixes between the Bloomington School of New Institutional Economics and the Virginia School of Political Economy are far from being exhausted. In what follows, we provide an overview of the importance of polycentricity and self-governance as part of Ostrom's "design principles"; provide a more rigorous model for the calculus of consent; show how to use the calculus of consent to better understand, and perhaps even measure, the degree of polycentricity and self-governance; and use this rigorous calculus of consent model to reconstruct Fuller's (1978) arguments about why certain issues are addressed by markets, others by democratic politics, others by courts, and others by hybrid intitutions.

In analyzing institutional change and the potential for political reform, one can follow a three-step procedure: (1) Estimate the optimal calculus of consent decision rule/s regarding the issue/s under consideration. This means that one should estimate how collectivized should decision-making be about that issue, and who should be part of the decision-making group. (2) Analyze the existing institutions, in order to understand the extent to which their de facto decision rules departs from optimal calculus of consent decision rules. This can be quite complicated, involving an analysis of institutional constraints and the nature of compound institutions. (3) Analyze the existing collective action mechanisms and the pressures created by institutional competition, in order to see to what extent a move closer to the calculus of consent optimum is likely to happen, and estimating the details of such a move (what is likely to change).

This paper develops the full theory regarding the first step, and illustrates this approach to institutional analysis with a variety of examples (thus, giving a glimpse of how to perform step 2). The more normative-tinted step 3 is outside the strict scope of this paper, but, nonetheless, one should bear in mind that the ultimate goal of economic analysis is to grasp the possibilities of institutional reform (Buchanan, 1964). The utility of institutional *analysis* ultimately depends on its usefulness for institutional *development*.

The next section takes a closer look at the connection between polycentricity and Ostrom's "design principles". Section 3 lays out the calculus of consent model and connects it to the IPF perspective. Section 4 illustrates the model with a wide range of examples. Section 5 distinguishes between natural and institutional polycentricity, and explains why natural polycentricity holds the key to understanding institutional diversity. It then shows that two of the calculus of consent parameters correspond to natural and institutional polycentricity, and provides the full theory of institutional diversity. Section 6 concludes by pointing to a variety of possible applications of this approach to institutional diversity, and highlighting some of the empirical challenges and difficulties.

# 2 "Design principles" for sustainable development

The complexity of most institutional arrangements and social phenomena raises often insurmountable difficulties with respect to the task of institutional reform and of managing or planning institutional transitions (E. Ostrom 2008). Consequently, institutional change necessarily involves the interplay between formal institutions and local culture, and a large element of unplanned evolution and bottom-up emergence (Boettke 1996; Boettke, Coyne & Leeson 2008). The question becomes, under what conditions can we reasonably expect a non-market spontaneous order to lead to long-term and sustainable growth? One of Elinor Ostrom and collaborators' major successes has been to address this question both theoretically and empirically. Ostrom's "design principles" offer a set of universal heuristics, applicable across a wide diversity of institutional arrangements. When the existing institutions comply with those "design principles", sustainable growth is more likely to occur. But the "design principles" are not defined very rigorously. Among them, the most vague and least well understood is the existence of a "certain level" of self-governance. Moreover, institutional evolution is said to be facilitated by an overall framework of "polycentricity". Yet, polycentricity is notoriously difficult to measure or evaluate. Before developing a more rigorous perspective on the meaning of self-governance and polycentricity, let us briefly review the "design principles" and their importance.

These "design principles derived from studies of long-enduring institutions for governing sustainable resources" (E. Ostrom 2005: p. 259) give us a broad perspective about institutional efficiency. These design principles do not point to a single set of universal rules, but are broad ideas mainly about how to design different kinds of systems of operational and collective choice rules under different contexts characterized by different sources of uncertainty. They are a set of institutional heuristics that have been shown empirically Ostrom (1990); ? and theoretically Wilson et al. (2013) to lead to sustainable development.

The design principles are (E. Ostrom, 1990: p. 90; 2005: p. 259; Wilson, E. Ostrom & Cox 2013): (1) Clearly defined boundaries for using scarce resources; (2) Proportionality between the benefits and costs of various actors; (3) Most individuals affected by the rules are included in the collective choice group that can modify these rules; (4) Monitors and enforcers of rules are accountable for their actions; (5) Graduated sanctions for breaking the rules; (6) Access to low-cost local arenas for conflict-resolution providing decisions perceived as fair; (7) External governmental authorities recognize, at least to some extent, the right to self-organize at the local levels; (8) "Appropriation, provision, monitoring, enforcement, conflict resolution, and governance activities are organized in multiple layers of nested enterprises" (E. Ostrom 2005: p. 259). The last two design principles refer to the ideas of self-governance and polycentricity. We can separate the arguments for polycentricity and self-governance in two sets: arguments from efficiency, and arguments from resilience. To some extent there is a trade-off between efficiency and resilience because optimizing rules based on past experience only accounts for the known hazards, creating increased vulnerabilities to new and unpredicted challenges (Carlson & Doyle 1999; 2000; 2002; Janssen, Anderies & E. Ostrom 2004; 2007; Janssen & Anderies 2013; Aligica & Tarko 2013).

In a nutshell, the argument from efficiency is that different common-pool resources are best managed at different scales. It is thus inefficient to have a governing body at a fixed geographic scale, and to assign to this governing body a large host of responsibilities about managing many different common-pool resources (E. Ostrom [1972] in McGinnis 1999; Ostrom, Bish & Ostrom, 1988; Oakerson, 1999; Aligica & Boettke, 2009; Aligica & Tarko, 2012). The alternative is polycentricity, which E. Ostrom (2005: p. 183) introduces in the following way: "By polycentric I mean a system where citizens are able to organize not just one but multiple governing authorities at differing scales. ... Each unit exercises considerable independence to make and enforce rules within a circumscribed domain of authority for a specified geographical area. In a polycentric system, some units are general-purpose governments while others may be highly specialized. .... In a polycentric system the users of each common-pool resource would have some authority to make at least some of the rules related to how that particular resource will be utilized."

The following analogy due to Walter Oi illustrates why polycentric organization tends to be more efficient than centralization. It used to be the case that only large pies were only sold, and "when all frozen pies had 12 inch diameters, apple was the most popular flavor – but when 7 inch pies came on the market, apple immediately fell to something like fifth place. [Oi's] explanation: When you're buying a 12 inch pie, the whole family has to agree on a flavor, and apple wins because it's everyone's second choice. With 7 inch pies, family members each get their pick, and almost nobody chooses apple." (Landsburg 2013). To the extent that eating your own preferred pie doesn't create significant negative externalities to other people, it is more efficient to have the smaller pies. For the exact same reason, to the extent that smaller scale jurisdictions don't create significant negative externalities to the other jurisdictions, it is more efficient to have institutional diversity rather than centralized consolidation. So, the polycentric system involves many numerous independent and often overlapping jurisdictions, but operating under an over-arching system of norms and/or rules, which manages the possible externalities and sets the framework for an overall cooperative arrangement between the jurisdictions (Aligica & Tarko 2012). Besides efficiency considerations, a polycentric system is also more resilient: "Because polycentric systems have overlapping units, information about what has worked well in one setting can be transmitted to others who may try it out in their settings. ... [W]hen small systems fail, there are larger systems to call upon – and vice-versa." (E. Ostrom 2005: p. 183). Thus, the polycentric organization creates the conditions for experimentations with different systems of rules, but, also, creates redundancies.

Such redundancies have sometimes been interpreted as signs of inefficiency and have been used as ar-

guments for administrative consolidation. In the key metropolitan debate, the claims of inefficiency have been contested (for a review see Aligica & Boettke 2009; Aligica & Tarko 2012; Boettke, Palagashvili & Lemke 2013), but even if one would accept some of them, the argument about resilience would still remain. A polycentric arrangement, as opposed to "institutional monocropping" (E. Ostrom, 2008), can be less efficient than the situation in which all communities implement the best possible system of rules. However, such a situation is often unrealistic for two reasons. (a) there simply is no universal set of rules that is optimal for all systems, as different societies often face different kinds of uncertainties and need to adapt their institutional structure to fit this diverse array of challenges (Carlson & Doyle 1999; Anderies, Jassen & E. Ostrom 2004; 2007; Aligica & Tarko 2013). (b) Even if or when such a universal system of optimal rules may exist, institutional monocropping would still be a vulnerable solution on the long term. As new challenges are always on the horizon, a large uniform system would be affected at a large scale by all these challenges. By contrast, in a polycentric system only some sub-systems are affected at one moment in time, and, thanks to redundancies, even those affected sub-systems can receive help from the others. Polycentricity thus contains the scale of unpredictable crises to more manageable levels. Although this may be inefficient over the short-term, it is more resilient over the long term.

Self-governance is the simple operational principle that creates the institutional diversity of a polycentric system. The "minimal recognition of rights to organize" leads to a situation in which a variety of approaches to similar problems are available at all times. This is the main way in which the limits of reason and foresight in regard to complex social issues can be overcome – by means of experimentation and imitation of success stories from one community to another. A variety of institutional systems is more likely to diminish the search time for effective rules, when abstract reason alone cannot possibly satisfactorily analyze the situation: variety is better at generating new institutional solutions to difficult challenges and, thus, it is more likely to enhance overall robustness. As Folke et al. (2002) put it, in the face of challenges, "resilient systems contain the components needed for renewal and reorganization", and the only way in which they are endowed with such a coping ability is to host "diversity – of species, of human opportunity, and of economic options".

Importantly, self-governance is useful not just as a mechanism for institutional discovery, but also for monitoring and enforcement purposes. Rules need enforcement in order to be "in-use" rather than merely "in-form", but in order for enforcement itself to be more than just "in-form", monitors also need monitoring. This seems to create a paradoxical infinite regress of monitors of monitors of monitors etc. Ostrom and collaborators' answer to this dilemma is not speculative or purely theoretical, but empirical. They have shown that self-governance alleviates this problem by creating a circle of rules rather than a linear hierarchy (Ostrom, 2005: p. 265). For instance, in a group of agents where the agents themselves take turns at being monitors, the problem is diminished to manageable levels because the monitor will now have a vested interest in making sure that the rules are followed and thus s/he will not need outside monitoring. Even if the monitors are hired from outside the community, and the complexity of the system thus increases by involving more people, the monitors would still have the desire to uphold the rules in order to satisfy the demand of those who pay them.

Systems that depart at great lengths from self-governance tend not only to lack legitimacy, but also to be rife with corruption, rent-seeking and wide-spread inefficiencies (Ostrom & Ostrom, 2004). It is always important to consider both the incentive structure and the knowledge available to those responsible for monitoring and enforcement, as well as for rule design in general. This "robust political economy" approach (Boettke & Leeson 2004; Leeson & Subrick 2006; Pennington 2011; Boettke 2012) leads to the conclusion that long-term sustainable development is promoted by an array of "design principles" that include selfgovernance and polycentricity.

# 3 The calculus of consent model as a tool for institutional analysis

The calculus of consent model (Buchanan and Tullock, 1999; Wagner, 1988) rests on the minimization of total expected cost of the collective decision-making system, T(n), which is defined as the sum of the external costs function, E(n), and decision costs function, D(n). The external costs reflect the expected total harm suffered by those who disagree with the collective decision but are forced to obey it, and the harm caused by mistaken decisions. The decision costs are all the procedural costs involved in setting up the collective decision-making process and in enforcing the decision. The variable  $n \in (0, 1]$  gives the range of possible "decision rules", i.e. it specifies the size of the majority required to pass a decision on the issue. For example, n = 50% is the majority rule, and n = 67% is the super-majority rule.

The "optimal decision rule",  $n^*$ , is the decision rule that minimizes total costs. The key idea for using the calculus of consent as a tool for institutional analysis is this: Depending on the issue under consideration, the external and decision costs can differ, and, hence, different issues have different optimal decision rules. For this reason, different issues are addressed by means of different kinds of institutions, which implement different de facto decision rules. For example, markets have very low decision rules, as a single person can often singlehandedly take the decision; democratic politics usually operates under the majority rule; regulatory agencies operate under lower-than-majority rules (the more independent they are, the lower the rule); juries operate under unanimity rule; etc. This point of view is similar to the one made by Djankov et al. (2003) that "[t]he shape of the IPF — and hence the efficient choice — varies across activities within a society, as well as across societies".

Buchanan & Tullock themselves have introduced a rudiment of this kind of institutional analysis (1999:

pp. 83-4 and chapter 5). In their example, they compare the costs of private action with the costs of "adverse collective decisions". Depending on which one is lower, and assuming a certain tendency towards efficiency, the issue is more likely to be addressed by private property institutions or by government action. More generally, however, we can compare the optimal rule for a variety of institutions, rather than just for private property and government action. This is exactly what this paper does.

The standard problem raised by the calculus of consent is determining the optimal decision rule, given the external and decision costs. But in applying the model to institutional analysis, we need to compare this hypothetical optimum with the de facto rule. The de facto decision rule may get to approximate the optimal rule by three different routes: (1) A direct route, in which the de facto institutions simply implement the optimal rule. For example, if the optimal rule is lower than 50%, logrolling may lower the de facto optimal rule. Similarly, constitutional changes, which can have very high external costs, are usually implemented with super-majorities in parliament (although by majority rule in popular referenda). (2) An indirect route, in which the composition of the group (who gets to be part of the decision group) is changed. Some of the examples below, such as federalism or private property, reflect this route. Technocracy and representative democracy, as opposed to direct democracy, are also important examples of this route. (3) A compound institutions route, in which several institutions, with different de facto decision rules, are involved in taking the decision, and the resulting compound decision rule is somewhere in between. For example, judicial review, involves compounding the majority rule of the legislative with the expert rule of the judiciary, leading to a higher than majority, but lower than unanimity, compound rule.

The calculus of consent model provides a general bird's eye view on institutional efficiency, and remains generally agnostic about the exact underlining behavioral details of the bargaining processes by which institutions are actually created or changed. Riker's (1962) theory of political coalitions, the theory of logrolling (Buchanan & Tullock 1962: chapters 9-11; Tullock, Sheldon & Brady 2002: chapter 3), various theories of political and social entrepreneurship (Wagner, 1966; Oakerson and Parks, 1988; Leeson and Boettke, 2009; Boettke and Coyne, 2009a,b), and E. Ostrom's "behavioral approach to the rational choice theory of collective action" (Ostrom, 1998) provide additional details into the microeconomics of institutional design. However, the broad equilibrium approach is sufficient for the present purposes. One can always go into deeper details, but, due to space constraints, some approximations must be made.

Efficiency can of course be understood normatively, but Buchanan has taken quite a strong stance in favor of interpreting efficiency as a purely descriptive tool for making predictions: "Appropriately thorough analysis should include an examination of the institutional structure itself in a predictive explanatory sense. ... [E]fficiency ... becomes a prediction of results that will tend to emerge from the exchange process, not a criterion for telling us what should be present in order to further some externally derived value norm." (Buchanan, 1968, pp. 5-6). This paper does not take a strong stance on this matter. If one is mainly concerned with prediction, one tends to use efficiency in the Buchanan-Stigler fashion. If one is mainly concerned with institutional development, one tends to use efficiency more as a normative benchmark (although not necessarily the only one) as in Ostrom's approach.

To make either specific predictions or specific normative evaluations, we need, first and foremost, to assume empirically valid functional forms for the external and decision costs functions. Buchanan & Tullock assume that external costs are a monotonically decreasing function of n, such as this:

$$E(n) = E_0 2^{-n/N_E}$$
(1)

Others have argued in favor of using U-shaped external costs, in order to account for the fact that rejecting efficient proposals also creates a cost — the larger the deciding majority, the larger the likelihood that it will fail to adopt some good measures (Guttman, 1998). Similarly, if one takes seriously the hold-out problem, one is led to an external cost function that rapidly increases in the range of large n (Robert Bish, personal communication). Finally, for a group of people who have very strong collectivist preferences (like those opting to live in a kibbutz), the external cost function may also be U-shaped because close to unanimous decisions would be disturbingly individualistic to them (Shruti Rajagopalan, personal communication). No such complications are considered here, but they may be important in actual applications. For what it's worth, in my view, such issues should be included in the decision costs, and not in the external costs, but this argument will have to wait for another day.

The decision costs are a monotonically increasing function of n:

$$D(n) = D_0 2^{n/N_D} \tag{2}$$

The calculus of consent parameters, whose numerical values differ from issue to issue, are as follows:  $E_0$  are the maximal external costs and  $D_0$  are the minimal decision costs (if the decision is dictatorially taken by a single person),  $N_E$  is the size of the decision group at which external costs become half of the  $E_0$  maximum, and  $N_D$  is the size of the decision group at which decision costs become double of the  $D_0$ minimum.  $E_0$  and  $D_0$  are level effects, while  $N_E$  and  $N_D$  are marginal effects.

The total cost is:<sup>1</sup>

$$T(n) = E_0 2^{-n/N_E} + D_0 2^{n/N_D}$$
(3)

<sup>&</sup>lt;sup>1</sup>If one prefers to define the parameters  $N_E$  and  $N_D$  not as half-size or double-size, but as some other multiples,  $m_E$  and  $m_D$ , the '2's in formula 3 are replaced by the desired multiples.

These formulae are a mathematical rendition reproducing (figures 2-4) the graphical illustrations provided by (Buchanan and Tullock, 1999, chapter 6). The insights from the next sections would be hard or impossible to reach without this explicit algebraic rendition of the original graphs because these insights result from exploring the meaning of the parameters in formula 3, and from being able to derive other, less intuitive, mathematical connections between the parameters.

The total expected cost, as defined by equation 3, is minimized for

$$n^* = -\frac{N_D N_E}{N_D + N_E} \frac{\log\left(\frac{D_0 N_E}{E_0 N_D}\right)}{\log 2} \tag{4}$$

The predictions of this calculus of consent model are the *patterns* that result from plotting the optimal rule  $n^*$  as a function of each parameter,  $D_0$ ,  $N_D$ ,  $E_0$ , and  $N_E$ , while keeping the other parameters constant (or conditional on their values). Section 5 provides an extensive analysis of the  $n^* = n^*(N_E)$  prediction, while section 6 briefly discusses some of the others.

The reason why different issues have different optimal decision rules is that they correspond to different values of the calculus of consent parameters,  $D_0$ ,  $N_D$ ,  $E_0$ , and  $N_E$ . To put it differently, each issue is located somewhere in this four-dimensional parameter space, and various regions of this space correspond to different kinds of possible institutional arrangements; i.e. for a given level of technological development, T, the set of all possible efficient (equilibrium) institutions is the field  $I_T(D_0, N_D, E_0, N_E)$ . Technological changes can both move an issue from one location in the parameter space to another location (i.e. an issue that was best addressed in one way in the past is now best addressed in a different way), and they can change the  $I_T$  field itself (i.e. the very nature of the efficient institutions available at each location in the parameter space can change).

We can connect this perspective to the IPF perspective (Djankov et al., 2003) in the following way. The optimal decision rule,  $n^*$ , gives the total loss minimization point on the IPF. We can map the shape of IPF by looking at the size of external and decision costs as the decision rule is moved either in the direction of more collective decision making ("state expropriation costs", E(n)) or more private decision making ("disorder", D(n)). The optimal levels of private disorder and collectivization are determined by  $D(n^*)$  and  $E(n^*)$ . Figure 1 illustrates the result by drawing the parametric plot from equations 1 and 2, thus reproducing in a mathematically rigurous way the intuitive IPF graphs drawn by Djankov et al. (2003). By doing this, we also obtain a more general perspective on the IPF concept, anchored in the Coasian logic of the calculus of consent.

To reiterate, the key idea for using the calculus of consent as a tool for institutional analysis is that, because different issues generate different external costs and may be subjected to differing decision costs,



Figure 1: The institutional possibilities frontier

different institutions are created to address them. Institutional efficiency can be increased (1) directly, by actually changing the decision rule, (2) indirectly, by changing the composition of the decision group, and (3) by compounding institutions. Let us consider several examples.

# 4 Examples of calculus of consent institutional analysis

I first look at examples, such as federalism and private property, when limiting the decision-making group makes sense because of low externalities on people outside this decision group. I then look at examples when it is actually efficient to have someone or a small group take decisions that affect everyone else. In cases such as calling the firemen anyone can be the self-appointed decider, while in many other cases there is a pre-designated group of "experts" taking the decisions for everyone else (e.g. representative democracy, regulatory agencies etc.). Furthermore, I look at checks on the possible problems created by such expert rule, i.e. at the creation of compound institutions.

#### Figure 2: Federalism

jority rule (very large external costs)





#### 4.1Federalism

Consider first the emergence of federal structures. What determines which issues are decided at the federal level and which are left at the local level? Tullock (1969) was the first to use a calculus of consent-like approach to model federalism based on the idea of optimal scales of governance. More recently, Inman and Rubinfeld (1997) explore different concepts of federalism based on the principle that one should "prefer the most decentralized structure of government capable of internalizing all economic externalities". Along similar lines, the calculus of consent gives us the following insight (Buchanan and Congleton, 1998, chapter 13): Considering that we are using the majority rule, rather than some arbitrary rule (i.e. the calculus of consent operates within the majority rule institutional constraint), if some issues were to be decided at the federal level, they would generate very high external costs (fig. 2a). As such, because the optimal rule is much higher than the majority rule, the political sector would not actually be able to muster a decision on that issue. Few politicians would venture to put forward any kind of proposal regarding such issues because, due to the high external costs, a decision at the federal level (in any direction) would alienate many of their voters. The institutional arrangement that gets around this problem is to let the issue be decided (in a variety of different ways) at the local levels. In other words, who belongs to the decision group is changed. When this happens, the external costs of the majority (at the local level) become lower (fig. 2b). This reflects V. Ostrom's point that "a republican remedy for the republican disease of majority tyranny is conceptualized by the compounding of the republic" (1987: p. 25). The example of federalism thus illustrates the indirect route by which the de facto rule gets closer to the optimal rule: letting the issue be decided at the local level means that many people are excluded from the decision group, but the key is to exclude only those who are not seriously affected by the decision.

As a more concrete example, consider the case of public libraries versus the inter-state highway system in the US. Why are libraries managed and financed at the local level? One way to understand this is to note the relatively large external cost of imposing the same set of books to every library across the country, or, alternatively, the large decision costs faced by a hypothetical benevolent central planner who tried to decide which books to send to each part of the country (i.e. a central planner that tried to replicate the optimal decision rule, resisting the temptation created by the fact that the decision could be taken much easier). By contrast, why is the inter-state highway system managed at federal level? This is because the network of local roads would not provide easy access for long-distance travel. The roads built for local needs are not the same as those required for long-distance travel. Hence, if the highway system was managed locally, each state would create large external costs to long distance travelers. The inter-state highway system addresses the demands of the entire country, without creating very large external costs on local communities.

It is important to point out that *actually reaching* the purely efficient institutions is not realistic, and we can only hope for *a tendency* towards increased institutional efficiency. Even by "compounding the republic", the decision might still not be fully optimal. For example, in fig. 2b the optimal rule at the local level is actually lower than the majority rule, so they now have the opposite problem than before: majority rule imposes decision costs that are a bit too high, and they might fail to reach agreement sometimes. Although there may be some tendency for the institutional system to get organized such that decision-making on various issues gets roughly allocated at the optimal levels, there always will exist room for improvement because the actual mechanisms for creating those improvements, such as institutional competition (Vanberg and Kerber, 1994; Caplan, 2001), public entrepreneurship (Wagner, 1966; Oakerson and Parks, 1988; Boettke and Coyne, 2009a), cultural contagion (Ostrom, 2008) or random shocks gradually filtering out unresilient systems (Wilson et al., 2013; Aligica and Tarko, 2013b), are far from perfect. This is why the Siglerian (1992) view is somewhat naive (institutional competition is not perfectly competitive).

### 4.2 Public and private goods

Why are some issues left to markets while others are dealt with by political means? We can answer this question by understanding private property as the limit case of federalism — the case in which the local level consists of just one person or very few people. So, private property is a bundle of issues for which anything less than unanimous consent would create significant external costs. For example, if we would decide by majority rule what everyone should eat or how to dress, the level of discontent would be quite high. In terms of the calculus of consent logic, imagine that the optimal rule in fig. 2a is very close to 100%. The federalism

solution discussed above is to restrict the composition of the decision group. In this case, the restriction goes down to all they way to the level of the individual or just a few individuals (clubs).

We can also think about what happens collectively once private property is in place. In other words, given private property, what is the incentive to move towards a more collective rule? A pure private good generates minimal external costs to other people. As such, the calculus of consent optimal rule tends to one person, i.e.  $n^* \rightarrow 0\%$ , and hence there is no incentive to try to replace private property with a more collectivist form of decision making. This simply means that I by myself am a big enough majority to take this decision optimally (in the social sense). To put it differently, the external costs of my decision are low enough that other people have no interest in coercing me in one direction or another. The subjectivity of these external costs is one of the critical issues addressed by the Bloomington School (Aligica and Tarko, 2013a) and it is a major subject in itself (Burnheim, 2006; Bozeman, 2002, 2007). At the other side of the spectrum, public goods involve high external costs (as judged by the people themselves). Collective decisions regarding an issue with much higher external costs, but in the same procedural environment as before (identical decision costs curve), will have a much higher equilibrium decision rule, e.g. in fig. 3b it is around 50%.

### 4.3 Emergency situations

While the example of federalism illustrates the case when the optimal rule is *higher* than the majority rule constraint, there are also cases when the optimal rule is *lower* than the majority rule constraint. The majority rule is now overbearing and the decision cannot be taken (in due time) by majority rule. However, it might be important to take a decision on the issue, and, if the increase in external costs caused by a lower-than-majority-rule decision is small, especially considering the costs of not acting, there is little benefit in sticking to the majority rule. Dougherty et al. (2014) show that "majority rule is typically more likely to select a Pareto optimal outcome than unanimity rule *for a infinite series of votes*" (emphasis added) because "although unanimity rule is better at retaining policies in the Pareto set ..., majority rule requires a lower threshold than unanimity rule to move policy into the Pareto set". By the same logic, an even lower decision rule discovers Pareto policies even faster — but also faces higher risks of moving out of the Pareto set.

A simple example along those lines is the fact that anyone can call the firemen. In other words, the optimal rule for such decisions, which are still collective decisions, is close to one single person. Using Guttman's (1998) perspective, we can also think of the costs of taking the wrong decision (e.g. calling the firemen when there's no fire) versus the costs of not taking the correct decision (e.g. failing to call the firemen when there is a fire). The costlier the false negative is relative to the false positive, the lower the optimal

#### Figure 3: Collective decisions



decision rule is.

## 4.4 Expert rule

In emergency situations, such as the one noted above, anyone can be the decider, but other cases, in which the optimal rule is lower than the majority rule, involve selecting a group of "experts" to take the decision for the group as a whole. Representative democracy itself is an example of such expert rule, in which direct democracy is replaced by indirect rule via representatives. In such cases, the question is how to select the exact composition of the small group that will take the decision for everyone (i.e. how experts or representatives are selected). Djankov et al. (2003) explore this issue by means of their IPF construct.

To better understand this issue, imagine a society of identical people, including having the same opinions regarding all collective decisions. In such a hypothetical case, any decision made by a single randomly selected person for the group as a whole would be optimal. Higher decision rules are necessary only because external costs are heterogeneous across different members of the group. As such, we can think of selecting experts as a method of reducing the decision costs created by heterogeneity. On the other hand, concerns with the self-interested or corrupt behavior of these experts, and concerns with the statistical representativeness of the expert group relative to the population as a whole, reflect concerns about the external costs of such a rule-by-expert system. Typical public choice considerations about rent-seeking and regulatory capture are included here in these external costs. They are details about what exactly is creating these external costs.

The calculus of consent logic of decision and external costs thus suggests that experts should be selected such that they have enough heterogeneity that the external costs of their decision are low (i.e. the heterogeneity of the expert group mimics the heterogeneity of the larger society), but is small enough that their decision costs are not too high. For example, this reasoning allows us to think systematically about the optimal size of parliaments (how many people per representative should there be). The reason why we have rule-by-experts, such as representative democracy, instead of direct democracy is that decision costs of direct democracy are often prohibitive. This also explains why direct democracy is more likely in smaller countries, and also why technological developments regarding mass-communication make direct democracy more feasible. The importance of heterogeneity is further explored in section 5.

Modern democratic states also incorporate and rely upon a variety of non-democratic institutions that are even more remote from voters than the representatives: independent regulatory agencies (IRAs). One can understand the calculus of consent rationale for such institutions in the following way. The high levels of expertise involved in various issues (such as monetary policy or various regulatory issues) create a high decision cost for a large group to zero-in on a single decision. Even assuming perfect Bayesian rationality, the more complicated the issue, the more time it takes for Bayesian updating to zero-in on a single point of view. Furthermore, all collective decision-making suffers from rational ignorance, because the more people are involved, the less incentive each of them has to get properly informed about the matter. This rational ignorance thus increases the decision costs even more. These are situations in which IRAs tend to be created or emergency powers are granted to various actors (fig. 4b).

Once again, although this argument shows that a *certain* level of expert-rule is efficient, it does not follow that the *actual* level of expert-rule is efficient. Matters of power, rent-seeking, voters' rational ignorance and irrationality create departures from the ideal (Buchanan et al., 1980; Acemoglu, 2003; Acemoglu and Robinson, 2006; Caplan, 2001; Acemoglu and Robinson, 2012; Caplan, 2007).

## 4.5 Compound institutions

An important point about representative democracies is that they are limited. As Caplan (2007: p. 190) pointed out: "Just because dictatorship is disastrous, it hardly follows that democracy must have free rein. Like markets, democracy can be limited, regulated, or overruled. Contramajoritarian procedures like judicial review can operate alongside democratic ones. Supermajority rules allow minorities to thwart the will of the majority." The question is, what determines the details of those limits? A previous example provided an explanation for why certain issues tend to be allocated by markets while others by political means. But what about *non-market* limitations of majoritarian democracy such as judicial review or decisions taken by IRAs? We've seen above why IRAs might be created, but IRAs are not actually entirely independent — they are subjected to political review, and to vagaries of politically-motivated appointments. Once again, the explanation hinges on the majority rule constraint. Such hybrid institutions like IRAs are created to get

around the fact that the majority rule is far from optimal with respect to certain issues; but, then, political control is implemented because technocratic decision-making with very low decision rules are not optimal either.

#### Figure 4: Compound institutions

(a) The case for judicial review (flat high decision costs (b) The incentive to create and politically control IRAs:
 make it possible to pass a decision by majority rule with If the decision were taken by majority rule, decision costs high externalities)
 would be much larger; but purely independent IRAs depart from public interest



The case illustrated in fig. 4a covers situations of judicial review. Here we have an example of increasing institutional efficiency by compounding institutions. Judicial review occurs when it is relatively easy to take a decision under majority rule, but the external costs might be large. This situation is illustrated by an almost flat decision costs function at a relatively high level. Because the decision costs are high even for low levels of n, the private property solution discussed above is not feasible, i.e. even if we restrict the composition of the decision group, the deciding group would still face very high decision costs. Consequently, the decision tends to be taken by collective procedures, if only because such procedures are already in place and adapted to cope with high decision costs.

Consider a few examples normally subjected to judicial review, such as a decision to grant a subsidy or to revoke a residence permit. Take first the decision costs. The private actions involved in trying to replicate a subsidy grant would be quite elaborate. They involve a charity-like enterprise for fund raising, in-kind contributions or time contributions, and a public campaign for creating social pressure against those who would not voluntarily contribute. Recycling or religion might provide such examples. Thus, the decision costs are high even if, strictly speaking, the size of the coalition deciding the issue is relatively small. Similarly, trying to replicate the revocation of a residence permit by purely private means is possible, but quite costly. It involves organizing a boycott campaign and/or a pressure campaign trying to determine the undesired party to sell and move. Consider now the external costs. The decision by majority rule may affect many or almost all people in the group. For example, many people may be taxed in order to provide for the subsidy, or many customers may be affected by the decision to revoke a firm's residence permit. Judicial review can be seen as acting as a check on the majority rule and a safeguard, accounting for the fact that, although the actual optimal rule is much higher than 50%, it is difficult to prevent the decision to be taken by majority rule. Basically, the decentralized solution is unfeasible because of the high flat decision costs function.

A similar type of argument supports the idea that IRAs should be subjected to a certain level of political control (fig. 4b). In this regard the discussion about them tends to be unnecessarily polarized. For example, a common question is, should monetary policy be done by an independent central bank or politically controlled by democratic representatives? Such a question is mistaken. The optimum is somewhere in between the majority rule and the low decision rule of a purely independent IRA. The correct calculus of consent question is *how much* independence should the IRA under consideration, e.g. the central bank, have? The situation in such cases is similar to the one discussed above regarding judicial review, but the roles of the decider and the monitor are reversed. In case of judicial review the decision is taken by democratic representatives, and the experts provide the evaluation and the veto. In case of IRAs, experts take the decisions while the democratic representatives do the evaluation and vetoing.

Consider the example of a central bank. The tasks of a central bank can be performed by private means (White, 1999). Before central banks, private banks organized into clearing houses, which were private associations designed to provide mutual monitoring and lender of last resort functions. To answer the question of whether the institutional change from the system of private clearing houses to the central banking system was a good idea, one needs to look at the external costs of the private clearing houses and the decision costs faced by politically controlled (quasi-independent) central banks. For example, Romer (1999) and Selgin et al. (2012) look at frequency of recessions, at the size of the losses during recessions, and inflation before and after the Federal Reserve was created. Levy and Peart (2010) look at the potential for regulatory capture faced by financial regulatory mechanisms. Such studies evaluate some of the key external costs of alternative banking arrangements. On the other hand, others look at the knowledge and incentives problems faced by actors under different banking arrangements (Salter, 2013). Such studies work towards evaluating the decision costs. We don't yet have good answers to the questions "How independent should a central bank be?" and, more radically, "Should we actually have central banks?" because the complete calculus of consent analysis for such a complex problem has not yet been properly performed, each study looking at only some piece of the overall puzzle.

## 4.6 Why the majority rule?

Some of the examples above provide an explanation for various institutions that get around the majority rule constraint. But isn't this begging the question? Why does the majority rule constraint exist at all? Buchanan & Congleton (1998: p. 18) point out that although "there is nothing either theoretically unique or morally sacrosanct about the majority decision rule, as such", nonetheless, in the exchange paradigm, the "majority rule may emerge as an important, perhaps the dominant, means through which political choices ... are made". The majority rule is a default equilibrium rule towards which a group of agents with heterogeneous and conflicting interests, and who are uncertain about which future policies might be attempted, gravitates. But, once this default rule is in place, and achieves its cultural mystique, the group must deal with special cases in which departures from the default are necessary — while at the same time preserving the special significance and importance of the 50% default. This idea is further elaborated by Guttman (1998) who points out that, when we combine the concern with efficiency with the concern for stability, the majority rule becomes optimal under a wide range of contexts. Thus, institutional diversity and complexity is explained by the fact that institutional solutions that often work fail under certain specific circumstances, and additional institutional devices are created for these specific circumstances, while, nonetheless, preserving the general broad institutions (such as the majority rule) that have wide-spread usefulness.

## 5 Polycentricity and decentralization

Polycentricity provides a subtle, yet simple, way of analyzing the institutional diversity described above. This section moves from the mere post-hoc calculus of consent rationalization of existing institutional arrangements to a more dynamic perspective, highlighting the incentives for institutional change. This works because all institutions have a certain *degree* of polycentricity, and, as I am about to show, polycentricity proves to be the key parameter allowing us to map the incentives for institutional change.

Polycentric organization is a system of multiple decision centers with overlapping jurisdictions, operating under an over-arching system of norms or rules (Ostrom, 1999, 2005; Wagner, 2005; Sproule-Jones et al., 2008; Aligica and Boettke, 2009; Aligica and Tarko, 2012). The existence of multiple decision centers is designed to create the conditions for self-governance. Polycentricity also creates the conditions for institutional competition, experimentation, and learning by doing. The overlapping of jurisdictions creates the ability to better spread knowledge, provide mutual assistance in cases of emergency, and enhance institutional competition by providing multiple choices to individuals without having them to move. But polycentricity is not anarchy. The over-arching system of norms and rules provides minimal universal standards, such as preventing "problems associated with local tyrannies and inappropriate discrimination" which "can be addressed by larger, general-purpose governmental units who are responsible for protecting the rights of all citizens and for the oversight of appropriate exercises of authority within smaller units of government" (E. Ostrom 2005: p. 183).

## 5.1 Natural and institutional polycentricity

Before going further, let us make an important distinction between two concepts of polycentricity, a distinction that will play a critical role in our later account. We can distinguish between natural polycentricity and institutional polycentricity. Polanyi (1951, 1962) and Fuller (1978) have mostly used the term in the natural sense, while the Bloomington School has mostly used the term in the institutional sense, using the idea of "optimal scale" to refer to natural polycentricity.

Natural polycentricity refers to the fact that certain issues are inherently polycentric, by their very nature. For example, Polanyi notes that the process of scientific discovery is inherently polycentric in the sense that one cannot discover truth unless one has competing perspectives working in conjunction, under a common set of norms about fraud, replicability, and information disclosure. Similarly, Fuller notes that certain problems brought before a court entail wide-spread consequences for society at large. Although there may be only two parties before the court, the decision can impact numerous *other* people and businesses. In such a case, the problem brough before the court is inherently polycentric. In Fuller's account, courts should not take decisions in polycentric matters because the court decision only takes into account the interests of the parties before the court and neglects broad externalities of the decision. Hence, the degree of polycentricity provides a criterion as to whether the issue is efficiently addressed by courts or by other institutions (markets and politics).<sup>2</sup>

Polanyi and Fuller use this concept of natural polycentricity to argue why certain forms of *institutional* polycentricity should be adopted. Thus, centralizing the scientific community is a bad idea because it would create a mismatch between the natural polycentricity of the scientific discovery process and its institutional organization. Similarly, in Fuller's account, issues that are naturally polycentric should not be addressed by courts, but by markets or democratic politics. In this way, either the diversity will be preserved (markets) or the wide-ranging conflicts, involving more than just the two parties before a court, would be addressed

<sup>&</sup>lt;sup>2</sup>Mirroring Buchanan's critique of democracy as science (Buchanan, 1967), Fuller acknowledges that conflict resolution requires institutions for bargaining and compromise, rather than for discovering some presumed impersonal truth, and he argues that court decisions are based on an assumption of truth-seeking. Consequently, if the issue involves inherent and widespread conflicts of interest, a court decision that tries to rely on some presumably universal and uniform set of reasons would be mostly delusional (about the absence of those inherent conflicts) and inefficient.

(politics).

The Bloomington School account of polycentricity is mostly focused on developing a clearer understanding of institutional polycentricity (Aligica & Boettke 2009; Aligica & Tarko 2012). The possibility of the mismatch between natural and institutional polycentricity has been discussed, especially with respect to identifying the kinds of issues with which a monocentric organization can actually deal effectively. For example, E. Ostrom (2005: pp. 278-9) notes that the US Geological Survey has provided a useful service in terms of information gathering, and that, due to the inherent large uncertainty of this information gathering process, the optimal scale of such an endeavor is very large.

In E. Ostrom's approach, the issue of natural polycentricity is thus quantified in terms of the concept of *optimal scale* for the provision of a public good or for managing a common-pool resource (see also Tullock 1969). The optimal scale is an objective feature of the issue at hand. The institutional structure is more or less efficient depending on how well the *de facto* scale of institutions fits the natural optimal scales of the issues they are addressing. However, as the examples from Polanyi and Fuller show, we need to understand this idea of optimal scale, not only in a geographical manner, but as referring to a network of affected agents. Many people in the same geographical area as the affected individuals or firms may not be affected at all, and the network of affected agents may be spread out across a very large geographical area. While in many cases geography provides a good proxi for this network, it does not always do so.

This distinction between natural and institutional polycentricity is already present in the mathematical version of the calculus of consent introduced earlier. Natural polycentricity is described by  $N_E$ , while institutional polycentricity can be identified with political decentralization,  $N_D$ . In the present model, institutional change is concerned with changing the value of  $N_D$ . The issue of informal leadership is briefly discussed in section 6. Leadership may be important in development economics contexts as an imperfect substitute for political decentralization and as a social-political mechanism for negociating the establishment of political units. In other words, we may have both formal and informal mechanisms for determining the actual value of  $N_D$ .

 $N_E$  is the speed with which external costs decline as the number of people involved in the decision increases, i.e. it is the marginal external costs of collective decision making. Because external costs are not institutionally contingent, they reflect natural, rather than institutional polycentricity. If external costs decrease very slow as we collectivize decision-making (i.e.  $N_E$  is large), it means that the issue is inherently (naturally) polycentric. By contrast,  $N_D$  is the speed with which decision costs increase as the number of people involved in the decision increases, i.e. it is the marginal decision costs of collective decision making. These decision costs depend on whatever institutional arrangements for social organization are in place.  $N_D$ thus reflects institutional polycentricity (i.e. political centralization and/or informal leadership). Consider the example of a politically decentralized country, in which individual members of parliament (who have to reach agreement on various collective decisions) represent local interests and have a large degree of independence (including financial independence if a large share of taxation is kept local). In such a country, the difficulty, i.e. decision costs D, of reaching a collective decision increases fast as the size of the deciding coalition, n, increases.  $N_D$  is the size of the coalition at which  $D(N_D) = 2D_0$ . The smaller  $N_D$  is, the smaller the marginal costs of adding another member to the deciding coalition is. If the country is politically decentralized, this marginal cost is going to be large because individual members can take a wider range of decisions without the consent of the collective. Informal leadership can be seen as another factor affecting  $N_D$  because the presence of effective leaders can facilitate consensus building, even when the formal rules are not inherently conducive to consensus.

# 5.2 The calculus of consent polycentric-based explanation of institutional diversity

To Fuller (1978), polycentricity provides the key to answering the question of why certain issues are addressed by markets, others by politics, others by courts, and others by hybrid institutions. He has struggled with some of the unintuitive aspects of this problem, providing some remarkable pieces of insight, but we can now re-create his argument in a much simpler and straightforward manner.

Figure 5: The optimal rule as a function of polycentricity, given the level of political centralization:

r

$$n^* = n^*_{N_D}(N_E)$$

Plotting the calculus of consent optimal rule as a function of natural polycentricity,  $n^* = n^*(N_E)$ , leads to a result that is, at least *prima facie*, unintuitive (fig. 5): as polycentricity increases, the optimal rule increases only up to a point, after which it starts declining. Plotting the result conditional on various levels of political decentralization,  $N_D$ , we see that, as one would expect, in highly centralized systems the optimal rule increases to high levels, while in highly decentralized systems the optimal rule remains relatively low (as people have cheap exit options). But why would the optimal rule as a function of polycentricy have a maximum?

This prediction makes perfect sense once we bear in mind that markets are themselves a typical example of a polycentric organization (Aligica & Tarko 2012): They consist of many independent decision centers operating under an over-arching system of rules (the definition and enforcement of property rights). Markets are thus an obvious institutional arrangement for addressing many issues that are inherently polycentric. Far from being unusual, the  $n^* = n^*(N_E)$  prediction actually provides the answer to Fuller's (1978) question about why certain issues are addressed by markets, others by democratic politics, and others by courts. As Fuller himself has argued, courts can only deal efficiently with issues that have low natural polycentricity. Beyond a certain increase of polycentricity, the issue should be addressed by either politics or markets. The calculus of consent  $n^* = n^*(N_E)$  prediction goes beyond Fuller (1978) by also answering the question regarding which high-polycentricy issues are addressed efficiently by markets and which by politics.

Markets, democracy, and courts. To illustrate, consider the simplest case of a majority rule democracy. Let us assume that issues that have optimal decision rules between 40 and 60 percent can be plausibly addressed by majoritarian politics — i.e. a decision, in one direction or another, can be passed. Depending on the natural polycentricity of the issue at hand, and the political centralization, its optimal rule may or may not enter the 40-60 percent band (fig. 6a). Sometimes the structure of the cost functions may be as such that for a particular set of issues there is no place for politics at all (fig. 6b).

Torts and criminal law. A longstanding question in law and economics has been why some issues are addressed by civil law while others by criminal law (Friedman 2000). Authors like Friedman (2000) have even questioned the need for criminal law altogether. The calculus of consent model can offer an explanation for the difference. The basic idea is that fig. 6a and 6b can be generated by a variety of combinations of the values of the other parameters, in particular for both low and high external costs,  $E_0$ . Acts covered by criminal law are generally seen as not merely acts against the directly injured party, but also as acts against society as a whole (or against the state governing society). This can be understood as meaning that the external costs are very large, going beyond the costs suffered by the party directly injured. For example, the social cost of murder involves general issues such people's expectation of safety. As such, if fig. 6a and 6b correspond to issues with high external costs, the courts described by the figure are criminal courts,





otherwise they are civil courts. When Friedman argues that criminal law could conceivably be replaced by tort law, he neglects the subjective external costs suffered by all those not directly injured by the crime.

**Hybrid institutions.** The present framework can also explain hybrid institutions. Rather than assuming a sharp transition from courts to politics to markets, as in fig. 6a and 6b, we can consider buffer zones covering these transitions. Hybrid institutions are created in those buffer zones (fig. 7a and 7b). Fig. 7a illustrates simple hybridity. Judicial review is in the buffer zone between courts and democratic politics, and IRAs are in the buffer zone between politics and markets.

Figure 7: The calculus of consent polycentric account of hybrid institutions,  $n^* = n^*(N_E)$ 





The situation may, however, be more complicated for issues with other kinds of cost curves. Institutional diversity inside the buffer zones is explained by the fact that the same relationship between the optimal rule and natural polycentricity can be obtained for a variety of values for the *other parameters*. For example, fig. 7b can be generated in a variety of ways, and, correspondingly, the buffer zone, B, reflects the need for

different kinds of hybrid institutions: if the figure is generated with a high value for  $N_D$  and  $D_0$ , reflecting a flat but high decision curve, the buffer zone corresponds to judicial review (as discussed earlier in section 4); if it is generated with a relatively small  $N_D$ , but a high  $D_0$  and  $E_0$ , it corresponds to independent regulatory agencies; if it is generated with small  $D_0$ ,  $E_0$  and  $N_D$ , it corresponds to private adjudication.

Political decentralization and federal oversight. Fig. 8a illustrates the tension between different administrative levels. As discussed in section 4, the optimal rule for some issues might be so large that a decision by majority rule would create disproportionately large costs to the losers as compared to the benefits for the winners. As such, when natural polycentricy is high and the optimal rule gets too high above 50%, there is an incentive to decentralize (i.e. increase the value of  $N_D$  — see fig. 5). Fig. 8a can thus be seen to describe an unstable institutional arrangement.

Figure 8: The calculus of consent polycentric account of federalism,  $n^* = n^*(N_E)$ 



To get a more detailed picture of the political controversy always surrounding the issue of political centralization or decentralization, consider the buffer zone D in fig. 8b. As the region on the right in fig. 8a, zone E in fig. 8b describes the incentive to decentralize. By contrast, the buffer zone D illustrates federal oversight, in line with E. Ostrom's concern about the need to address "problems associated with local tyrannies and inappropriate discrimination" (2005: p. 183). When the level of natural polycentricity is relatively high, but the optimal decision rule is still not very far from 50%, we always witness this debate about whether the issue at hand is important enough that a single uniform approach should be imposed on everyone or, on the contrary, a diversity of approaches may be permitted. The buffer zone D in fig. 8b illustrates this unavoidable area of controversy.

# 6 Empirical applications: The socio-political meaning of calculus of consent parameters

The calculus of consent predictions are derived by using eq. 4, and they are illustrated by plotting the optimal decision rule  $n^*$  as a function of each parameter, holding the other parameters constant. It is this step which would not have been possible without the algebraic rendition of Buchanan & Tullock's graphs. Section 5 has looked more thoroughly at  $n^* = n^*(N_E)$ , but many other predictions follow from eq. 4, and they are too numerous to be explored in a single paper. Those predictions are broad *patterns*, rather than specific expectations about specific cases.

More specific empirical applications rely on interpreting the parameters  $(D_0, N_D, E_0, N_E)$  and better pinpointing methods for their operationalization and measurement. To conclude the paper, this section provides a quick tentative description, highlighting some of the possibilities. The purpose here is not to solve any of those thorny empirical issues, but only to highlight that the model developed here is not purely theoretical, but can be potentially connected to a variety of important empirical problems.

## 6.1 Identifying the de facto decision rule

When trying to measure the external costs, E, we face the standard utility aggregation problem. We have to face the fact that the harm caused by the collective decision to each individual subjected to that decision has to be considered as subjectively perceived by each of them (Aligica & Tarko 2013a). We may, nonetheless, approximately solve this problem in the typical Kaldor-Hicks manner by looking at the resources spent by each individual (or, more realistically, by pressure groups) in the attempt to influence the decision (Becker, 1983; Wittman, 1995, chapter 7). At least approximately, the bigger the share of one's income (including the monetary value of time) spent in such a manner, the more important the issue is to that person. This way of aggregation leads to the intuitive consequence that more overall resources are spent trying to influence important collective decisions as compared to unimportant collective decisions. Thus, the total amount of resources spent in the attempt to influence the decision provides a rough measure of the maximal external costs,  $E_0$ , while the *actual* external costs,  $E(\hat{n})$ , given the *de facto* decision rule,  $\hat{n}$ , are the share of those total resources spent by the losers. If we also manage to measure  $N_E$  by some method, we can use this approach to calculate the de facto decision rule,  $\hat{n}$ , and compare it to the hypothetical optimum, hence, empirically estimating institutional efficiency.

Section 5 has argued that  $N_E$  should be interpreted as a measure of polycentricity, and different institutions are created to address issues with different degrees of "natural polycentricity". As such, each type of institution corresponds to a particular range of values for  $N_E$ . Looking at which specific institution pressure groups are trying to influence, one can get a rough estimate of  $N_E$  — hence, one can estimate all the parameters required to measure the *de facto* decision rule and, thus, institutional efficiency. The extent to which this can be done with sufficient precision is a challenge to the ingenuity of empirical researchers.

## 6.2 Heterogeneity, civic capital, and state capacity

Figure 9: The impact of social cohesion and of the importance of the issue on the optimal rule



We can also look at the matter through a representative agent lens. The maximal external costs  $E_0$  can be expanded as

$$E_0 = H(1-h)$$

where H is the importance of the issue to the representative loser, i.e. the maximal harm caused by the decision to the representative agent if it disagrees with the enforced decision (this is possible if a certain level of authoritarianism exists); while h is a measure of social cohesion or homogeneity (Easterly et al., 2006; Hulse and Stone, 2007; Heller, 2009; Janmaat, 2011). The impact of social cohesion on maximal external costs is due to the fact that when the members of the group agree, fewer of them are required for the decision. This can be seen as the statistical foundation of representative democracy: if there are significant clusters of views within the population, selecting a few people from each homogenous group to represent the group will not cause significant external costs. Fig. 9a and 9b illustrate how social cohesion and the importance of the issue impact the optimal decision rule for that issue.

The minimal decision costs  $D_0$  can be understood as a measure of state weakness. The higher the state capacity (Besley and Persson, 2008; Persson and Besley, 2009; Besley and Persson, 2010), the cheaper it is to take and enforce a collective decision (state capacity may be different with respect to different issues).

Failed states have high decision costs for many or all collective issues. This being said, as mentioned earlier, informal leadership can be a substitute to formal political centralization. Hence, one path away from failed states involves local leaders building larger scale consensus about what rules and institutions to put in place (Boettke et al., 2008; Coyne, 2008; Leeson, 2008a; Leeson, 2008b).

This can also be seen from the IPF perspective, as increases of either  $E_0$  or  $D_0$  lead to a move outward of the IPF (figure 1). This is either the result of higher "civic capital" (as in Djankov et al. (2003) interpretation) or of higher state capacity. One way to understand this is to think of power as a substitute for self-governance. This allows us to understand some of the difficulties involved in creating stable social orders. For example, discussing the problem of externally imposed order, Coyne (2008) notes that "a successful reconstruction requires the continued enforcement of property rights in the long term, after the occupiers exit" as "indigenous parties must be convinced that agreements reached during the occupation will be binding and enforceable after the occupation ends" (p. 66). In other words, while the occupiers are present, order is maintained by means of power, but, as they leave, if self-governing mechanisms are not in place to substitute for that power, order breaks down.

## 6.3 A self-governance perspective on political and economic freedom

The calculus of consent model also leads to a more rigorous perspective on self-governance. Self-governance simply means that the *de facto* decision rule,  $\hat{n}$ , describing the operation of the actual institutions, is close to the calculus of consent *optimal* decision rule,  $n^*$ . Dahl (1989) gets at this point intuitively when he notes that "to live in association with others necessarily requires that they must sometimes obey collective decisions that are binding on all members of the association. The problem, then, is to discover a way by which the members of an association may make decisions binding on all and still govern themselves." (p. 89). Similarly, Buchanan wrote about the "paradox of 'being governed'" and pointed out that "the individual does not enter into social contract [with others] for the purpose of imposing constraints on himself", but in order to "secure the benefits of behavioral limitation on their part" (Buchanan, 1975, p. 136).

As shown by the examples from section 4, self-governance may be achieved either directly, by having the de facto decision rule approximate the optimal rule, as in the case of private property rules or supermajority requirements for parliaments deciding constitutional changes, or indirectly, by having additional institutions (such as federalism, independent regulatory agencies or judicial review) mediate and go around the institutional constraint created by other institutions such as the majority rule constraint. Logrolling, as analyzed by Buchanan & Tullock (1999: chapters 9-11), and broader attempts at coalition-building (Riker 1962), can also be seen as indirect mechanisms for changing the de facto decision rules for a set of issues, despite preserving the majority rule in form.

We can define the *percent departures from self-governance* for a particular issue as:

$$\gamma(\hat{n}) = \frac{\hat{n} - n^*}{\hat{n} + n^*} \in [-1, 1]$$
(5)

There are two different types of departures from ideal self-governance, corresponding to the positive or negative values of  $\gamma$ . One type of self-governance imperfection is authoritarianism, i.e. lack of political freedom, when  $\gamma$  is negative. The other one is the tyranny of majority, when  $\gamma$  is positive. We can thus see the two main concerns with freedom as being two different facets of the same general issue. The *costs* of departing from self-governance are captured by the IPF graph (fig. 1).

Authoritarianism corresponds to cases when the optimal rule is higher than the actual rule, i.e. decisions affecting the entire collective are taken by a too-small subset of that collective. Empirical measures such as the Freedom House index, the Polity Score or The Economist's Democracy Index can be seen as intuitively trying to measure departures of  $\gamma$  from zero in the negative direction. The present conceptualization may provide clues for improving such measures (see Munck and Verkuilen (2002) for a critique of existing measures). The critical issue neglected by such indices is that of the optimal scale of governance (Tullock 1969; V. Ostrom 1987) and polycentricity.

The tyranny of the majority corresponds to cases when the optimal rule is lower than the actual rule, i.e. decisions that would optimally be taken by individuals or small groups are instead taken by larger collectivities. As Caplan (2007: p. 192) pointed out, "[u]ndemocratic politics is not the only alternative to democratic politics. Many areas of life stand outside the realm of politics, of 'collective choice'. When the law is silent, decisions are 'up to the individual' or 'left to the market'." Empirical measures such as the Fraser Institute's Economic Freedom of the World index intuitively try to measure departures of  $\gamma$  from zero in the positive direction. Once again, the present conceptualization may help improve the design of such measures, for example alleviating the suspicion of libertarian bias.

It is beyond the scope of this paper to do a thorough critique of such political and economic freedom measures. It is enough to point out here the broad principle, captured by eq. 5, about how to approach this issue of self-governance through the general perspective of the calculus of consent.

# 7 Acknowledgments

I am grateful to Robert Bish, Peter Boettke, Christopher Coyne, Randall Holcombe and Virgil Storr for very useful discussions based on previous versions of this paper. Previous versions were presented at the 2014 *Public Choice Society* meeting, and at the Spring 2014 Graduate Student Paper Workshop at the *Mercatus Center.* I thank all the participants for their comments.

# References

- Acemoglu, D. (2003). Why not a political Coase theorem? Social conflict, commitment, and politics. *Journal* of Comparative Economics 31, 620–652.
- Acemoglu, D. and J. A. Robinson (2006). Economic Origins of Dictatorship and Democracy. Cambridge: Cambridge University Press.
- Acemoglu, D. and J. A. Robinson (2012). Why Nations Fail: The Origins of Power, Prosperity, and Poverty. New York: Crown Publishing.
- Aligica, P. D. (2014). Institutional Diversity and Political Economy: The Ostroms and Beyond. Oxford: Oxford University Press.
- Aligica, P. D. and P. Boettke (2009). Challenging Institutional Analysis and Development: The Bloomington School. Routledge.
- Aligica, P. D. and V. Tarko (2012). Polycentricity: From Polanyi to Ostrom, and beyond. Governance 25(2), 237–262.
- Aligica, P. D. and V. Tarko (2013a). Co-production, polycentricity and value heterogeneity: The Ostroms<sup>2</sup> public choice institutionalism revisited. *The American Political Science Review 107*, 726–74.
- Aligica, P. D. and V. Tarko (2013b). Institutional resilience and economic systems: Lessons from Elinor Ostrom's work. *Comparative Economic Studies*, doi:10.1057/ces.2013.29.
- Arena, P. (2010). Bargaining models of war. Department of Political Science, University at Buffalo (SUNY).
- Becker, G. S. (1983). A theory of competition among pressure groups for political influence. The Quarterly Journal of Economics 98(3), 371–400.
- Besley, T. (2007). The new political economy. The Economic Journal 117, F570–F587.
- Besley, T. and T. Persson (2008). Wars and state capacity. Journal of the European Economic Association 6(2-3), 522–530.
- Besley, T. and T. Persson (2010). State capacity, conflict, and development. Econometrica 78(1), 1–34.

- Boettke, P. J. (2012). Living Economics: Yesterday, Today, and Tomorrow. Oakland, CA: The Independent Institute.
- Boettke, P. J. and C. J. Coyne (2009a). Context matters: Institutions and entrepreneurship. *Foundations* and Trends in Entrepreneurship 5, 135–209.
- Boettke, P. J. and C. J. Coyne (2009b). An entrepreneurial theory of social and cultural change. In Markets and civil society: the European experience in comparative perspective, pp. 77–103. New York: Berghahn Books.
- Boettke, P. J., C. J. Coyne, and P. T. Leeson (2008). Institutional stickiness and the new development economics. *American Journal of Economics and Sociology* 67, 331–358.
- Bozeman, B. (2002). Public value failure: When efficient markets may not do. Public Administration Review 62, 145–161.
- Bozeman, B. (2007). Public Values and Public Interest: Counterbalancing Economic Individualism. Washington, D.C.: Georgetown University Press.
- Buchanan, J. M. (1964). What should economists do? Southern Economic Journal, 213–222.
- Buchanan, J. M. (1967). Politics and science: Reflections on Knight's critique of Polany. Ethics 77, 303–310.
- Buchanan, J. M. (1975). The Limits of Liberty: Between Anarchy and Leviathan. Indianapolis: Liberty Fund, Inc.
- Buchanan, J. M. (1987). The constitution of economic policy. American Economic Review 77, 243-50.
- Buchanan, J. M. (1999 [1968]). The Demand and Supply of Public Goods. Indianapolis, IN: Liberty Fund, Inc.
- Buchanan, J. M. and R. D. Congleton (1998). Politics by Principle, Not Interest. Cambridge, UK: Cambridge University Press.
- Buchanan, J. M., R. D. Tollison, and G. Tullock (Eds.) (1980). Toward a Theory of the Rent-Seeking Society. Texas A&M University Press.
- Buchanan, J. M. and G. Tullock (1962 [1999]). The Calculus of Consent. Liberty Fund.

Burnheim, J. (2006). Is Democracy Possible? Sydney: Sydney University Press.

- Caplan, B. (2001). Standing tiebout on his head: Tax capitalization and the monopoly power of local governments. *Public Choice 108*, 101–122.
- Caplan, B. (2007). The Myth of the Rational Voter: Why Democracies Choose Bad Policies. Princeton University Press; New edition.
- Coyne, C. J. (2008). After War: The Political Economy of Exporting Democracy. Stanford, California: Stanford University Press.
- Dahl, R. A. (1989). Democracy and Its Critics. Yale University Press.
- Djankov, S., E. Glaeser, R. La Porta, F. Lopez-de Silanes, and A. Shleifer (2003). The new comparative economics. *Journal of comparative economics* 31(4), 595–619.
- Dougherty, K., B. Pitts, J. Moeller, and R. Ragan (2014). An experimental study of the efficiency of unanimity rule and majority rule. *Public Choice* 158(3-4), 359–382.
- Easterly, W., J. Ritzen, and M. Woolcock (2006). Social cohesion, institutions, and growth. *Economics & Politics 18*, 103–120.
- Fearon, J. D. (1995). Rationalist explanations for war. International Organization 49, 379-414.
- Filson, D. and S. Werner (2002). A bargaining model of war and peace: Anticipating the onset, duration, and outcome of war. American Journal of Political Science 46, 819–837.
- Friedman, D. D. (2000). Law's Order: What Economics Has to Do with Law and Why It Matters. Princeton, NJ: Princeton University Press.
- Fuller, L. (1978). The forms and limits of adjudication. Harvard Law Review 92, 353-409.
- Guttman, J. M. (1998). Unanimity and majority rule: the calculus of consent reconsidered. European Journal of Political Economy 14, 189–207.
- Heller, L. (2009). Social cohesion and its relationship to endogenous institutional formation and economic growth. Journal of Private Enterprise 25, 81–104.
- Hirshleifer, J. (2001). The Dark Side of the Force: Economic Foundations of Conflict Theory. Cambridge, UK: Cambridge University Press.
- Hulse, K. and W. Stone (2007). Social cohesion, social capital and social exclusion. *Policy Studies 28*, 109–28.

- Inman, R. P. and D. L. Rubinfeld (1997). Rethinking federalism. *The Journal of Economic Perspectives*, 43–64.
- Janmaat, J. G. (2011). Social cohesion as a real-life phenomenon: Assessing the explanatory power of the universalist and particularist perspectives. Social Indicators Research 100, 61–83.
- Leeson, P. T. (2006). Cooperation and conflict. American Journal of Economics and Sociology 65(4), 891–907.
- Leeson, P. T. (2008a). Coordination without command: Stretching the scope of spontaneous order. Public Choice 135, 67–78.
- Leeson, P. T. (2008b). How important is state enforcement for trade? American Law and Economics Review 10, 61–89.
- Leeson, P. T. and P. J. Boettke (2009). Two-tiered entrepreneurship and economic development. International Review of Law and Economics 29, 252–259.
- Leeson, P. T., C. J. Coyne, and P. J. Boettke (2006). Converting social conflict: Focal points and the evolution of cooperation. *The Review of Austrian Economics* 19(2-3), 137–147.
- Levy, D. M. and S. J. Peart (2010). Prudence with biased experts: Rating agencies and regulators. *Working* paper.
- McGinnis, M. D. (Ed.) (1999a). Polycentric Governance and Development: Readings from the Workshop in Political Theory and Policy Analysis. University of Michigan Press.
- McGinnis, M. D. (Ed.) (1999b). Polycentricity and Local Public Economies: Readings from the Workshop in Political Theory and Policy Analysis. University of Michigan Press.
- Munck, G. L. and J. Verkuilen (2002). Conceptualizing and measuring democracy: Evaluating alternative indices. Comparative Political Studies 35, 5–34.
- Nye, J. V. C. (1997). Thinking about the state: Property rights, trade, and changing contractual arrangements in a world with coercion. In *Frontiers of the New Institutional Economics*. Academic Press.
- Oakerson, R. J. and R. B. Parks (1988). Citizen voice and public entrepreneurship: The organizational dynamic of a complex metropolitan county. *Publius: The Journal of Federalism 18*, 91–112.
- Ostrom, E. (1990). Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge University Press.

- Ostrom, E. (1998). A behavioral approach to the rational choice theory of collective action. *The American Political Science Review 92*, 1–22.
- Ostrom, E. (2005). Understanding Institutional Diversity. Princeton: Princeton University Press.
- Ostrom, E. (2008). Developing a method for analyzing institutional change. In Alternative Institutional Structures: Evolution and Impact. Routledge.
- Ostrom, E., R. Gardner, and J. Walker (1994). Rules, Games, and Common-Pool Resources. University of Michigan Press.
- Ostrom, E. and V. Ostrom (2004). The quest for meaning in public choice. *American Journal of Economics* and Sociology 63, 105–47.
- Ostrom, V. (1987). The Political Theory of a Compound Republic. Lincoln and London: University of Nebraska Press.
- Ostrom, V. (1997). The Meaning of Democracy and the Vulnerability of Democracies: A Response to Tocqueville's Challenge. Michigan: University of Michigan Press.
- Ostrom, V. (1999). Polycentricity (part 1 and 2). In *Polycentricity and Local Public Economies*, pp. 52–74 and 119–138. Michigan: University of Michigan Press.
- Persson, T. and T. Besley (2009). The origins of state capacity: Property rights, taxation and policy. *The* American Economic Review 99(4), 1218–1244.
- Polanyi, M. (1951). The Logic of Liberty. University of Chicago Press.
- Polanyi, M. (1962). The republic of science: Its political and economic theory. Minerva 1, 54-74.
- Powell, B. and E. P. Stringham (2009). Public choice and the economic analysis of anarchy: A survey. *Public Choice* 140, 503–538.
- Reiter, D. (2003). Exploring the bargaining model of war. Perspectives on Politics 1, 27–43.
- Romer, C. D. (1999). Changes in business cycles: Evidence and explanations. Journal of Economic Perspectives 13(2), 23–44.
- Salter, A. (2013). Robust political economy and the lender of last resort. *George Mason University, Economics Department.*

- Selgin, G., W. D. Lastrapes, and L. H. White (2012). Has the Fed been a failure? Journal of Macroeconomics 34(3), 569–596.
- Smith, A. and A. C. Stam (2004). Bargaining and the nature of war. *Journal of Conflict Resolution* 48, 783–813.
- Sproule-Jones, M., B. Allen, and F. Sabetti (Eds.) (2008). The Struggle to Constitute and Sustain Productive Orders: Vincent Ostrom's Quest to Understand Human Affairs. Plymouth, UK: Lexington Books.
- Stigler, G. J. (1992). Law or economics? Journal of Law and Economics 35, 455-468.
- Tullock, G. (1969). Federalism: Problems of scale. Public Choice 6(1), 19–29.
- Tullock, G. (2005 [1974]). The Social Dilemma of Autocracy, Revolution, Coup D'Etat, and War. Indianapolis, Indiana: Liberty Fund.
- Vanberg, V. and W. Kerber (1994). Institutional competition among jurisdictions: An evolutionary approach. Constitutional Political Economy 5, 193–219.
- Wagner, R. E. (1966). Pressure groups and political entrepreneurs: A review article. Papers on Non-Market Decision Making 1, 161–170.
- Wagner, R. E. (1988). The 'Calculus of Consent': A Wicksellian retrospective. Public Choice 56, 153–66.
- Wagner, R. E. (2005). Self-governance, polycentrism, and federalism: Recurring themes in Vincent Ostrom's scholarly oeuvre. Journal of Economic Behavior & Organization 57, 173–188.
- White, L. H. (1999). The theory of monetary institutions. Blackwell Oxford.
- Wilson, D. S., E. Ostrom, and M. Cox (2013). Generalizing the core design principles for the efficacy of groups. Journal of Economic Behavior & Organization 90, S21–S32.
- Wittman, D. A. (1995). The myth of democratic failure: Why political institutions are efficient. University of Chicago Press.