Self-Enforcing Federalism: Solving The Two Fundamental Dilemmas

Rui J. de Figueiredo, Jr., and Barry R. Weingast

April 1997

1. Introduction

Although federal systems differ on many dimensions, all face the *two fundamental dilemmas of federalism*:

> **Dilemma 1:** What prevents the national government from destroying federalism by overwhelming its constituent units?
>
> **Dilemma 2:** What prevents the constituent units from undermining federalism by free-riding and other forms of failure to cooperate?

To survive, a federal system must resolve both dilemmas. This requires that the rules defining federalism be self-enforcing for political officials at all levels of government. A theory of the performance of federalism must therefore analyze the incentives of political officials to abide by the rules.

Resolving the two dilemmas is problematic because they imply a *fundamental tradeoff*: solving one dilemma exacerbates the other. Placing stronger constraints on the national government in an effort to mitigate the first dilemma usually weakens the constraints on the constituents units, exacerbating the second dilemma. Stronger constraints on the constituent units, usually imposed by the national government, risks exacerbating the first dilemma.

The survival of federalism therefore requires a delicate balance between these two problems. A national government that is too weak will exhibit free-riding and insulated, "dukedom" economies. Or worse, it will disintegrate. With a national government too strong, federalism typically fails because the national government compromises state independence, extracting rents from the states and hindering or prohibiting interstate competition that underpins the positive economic effects of federalism. Reflecting this tradeoff, several theorists remark about federalism's instability (Riker 1964, Bednar 1997).

Unfortunately, most studies of federalism take its structure as given. Economists, for example, have long-studied the assignment problem concerning the optimal allocation of taxation and policy authority across levels of government (Oates 1972, Rubinfeld 1988), but they rarely study how actual federalisms create rules allocating these powers that can be sustained. Most political scientists also take federal structures as given. An important exception is Riker (1964).

In this paper, we develop a model showing how the two dilemmas operate *simultaneously*. We present a repeated game that captures the nature of federal arrangements. By endogenizing federal power, state participation and shirking, and limits on the federal government, we derive a set of sufficient conditions for a stable federal system. Further, the model affords results about the welfare characteristics of federal systems.

Our work contributes to a new and growing literature which Gibbons and Rutten (1996) call the new "equilibrium institutionalists." Scholars in this new tradition observe that any constitutional feature must be sustained by political actors with an incentive to abide it. All government institutions — such as democratic elections, separation of powers, federalism — and all citizen rights — such as the right to vote, to own property, to form free associations, to free expression — impose limits on government officials. For these rights to exist in practice, officials must have incentives to honor them..[1]

Several authors have begun to study the two fundamental dilemmas of federalism from this perspective. Bednar (1996), defining the "N+1 game," and Treisman (1997) use different

---

[1] Major contributions include Bednar (1996), Calvert (1992,1996), Gibbons and Rutten (1996), Greif, Milgrom, and Weingast (1994), Milgrom, North, and Weingast (1989), and Weingast (1997).

approaches to study the second dilemma. Weingast (1995) studies the first dilemma. Ordeshook (1996) sketches an "N+1" game to study federalism focusing on the role of political parties. No scholar studies both fundamental dilemmas simultaneously.

To understand how successful federal systems resolve the two dilemmas, and thus provide for their stability, we begin with the reasons for attempts to construct federal systems. Broadly speaking, federalisms are motivated by opportunities to capture gains from hierarchy. The case of an agglomeration of independent states, called *bottom-up* federalism, is motivated by opportunities to capture gains from exchange. Federal systems can also be promulgated by a single unit; these cases of *top-down* federalism are typically motivated by the central unit's desire to reap gains from specialization and decentralization.

The first question about bottom-up federalism concerns why these systems need a central structure at all. The answer lies in the fact that many policies produce a conflict between individual and social rationality. Although participating states want to enjoy the benefits of public goods provision, each has an incentive to *shirk*, to let the others contribute while they "free-ride" (Bednar 1996). In some cases, such as small numbers of states with large benefits from trade, decentralized cooperation is possible through repeated play, without a central structure; participating states can induce potential defectors to contribute by threatening to expel them from the federation (Axelrod (1984); Kreps and Wilson (1990)). But if the states are large in number, or the benefits are not large in relation to costs, a decentralized solution may not be possible. Imperfect information about who contributes exacerbates these problems, since it is harder to sanction states if others cannot identify those that shirk (Green and Porter (1984); Persson and Tabellini (1994); Bednar (1996); Milgrom, North, and Weingast (1990)). In these cases, hierarchy and centralization potentially allow the realization of the benefits of cooperation. When decentralization fails, federalism is a possible solution.

Bottom-up federalism can perform two functions. First, the central government can be an *agent* of the states, to which they delegate responsibility for providing public goods. Second, federalism, by centralizing and diffusing information, can enhance reputation mechanisms if moral hazard problems prohibit the successful implementation of decentralized

punishment strategies (e.g. Milgrom, North and Weingast (1990); Green and Porter (1984)). In this paper, we confine our discussion to the first type of functions.

Central governments frequently do not possess resources of their own. When a set of independent states create federalism to provide public goods, states must contribute the resources needed by the central government to produce public goods. National defense provides an example. In some instances, a set of independent states can provide for their own defense. Because of a commonality of security interests and large scale effects, however, it is often beneficial for the states to cooperate on defense; creating a national authority which provides this good is an attractive alternative. But to erect such a defense, the central authority requires contributions from the states.

This case highlights one of the primary problems of federalism. Because the goods are not (fully) excludable, each state has an incentive to free-ride on the contributions of other states. This *moral hazard* problem is compounded if the individual contributions are not fully observable. A primary solution, which reinforces the justification for erecting the national authority, is to provide the center with policing authority; it acts as a central monitor in the hierarchical structure. Thus, one of the key design choices in federalism is the grant of power for provision of public goods and policing states.

If the central government is a faithful agent of the states, then federalism poses no design puzzles. In particular, states would grant as many resources as the federal government needed for the optimal provision of public goods and to prevent shirking. National governments are not without their own interests, however. As noted earlier, in granting resources and powers to the central government, the states also enable it to usurp state authority and extract resources. Indeed, the more institutional and economic power the center has to carry out its delegated tasks, the greater will be the potential for *encroachment* on state sovereignty and authority.

These two phenomena, reflecting the fundamental tradeoff noted above, represent the central design puzzle of federalism: states must grant substantial power and resources to the center to prevent moral hazard by the states and provide sufficient public goods; but these resources increase the central government's ability to encroach on the states. The example of

defense makes clear the tradeoff: giving the national government greater resources allows appropriate defense against external threats to the states; but increasing central resources also makes it harder for the states to resist encroachments by the center, making it more likely that the national government will usurp state authority. An increase in institutional power garners benefits but at the cost of sovereignty and potential benefits. If the choice of institutional power for all levels of government is not *self-enforcing*, the federalism will ultimately fail.

In this paper, we do not attempt a comprehensive model of all classes of federalisms. Instead, we choose a subset of these problems and illustrate the challenges and solutions. In Section 2, we consider a decentralized case of N states without a hierarchy. Here we show that, under certain circumstances, if there are gains from exchange, a decentralized structure may capture those benefits. However, if there are a large number of states, imperfect information, low benefit to cost ratios or low discount factors, such decentralized solutions will unravel.

In Section 3, we consider a base-case centralized structure, in which federal power to provide goods is *not* tied to its ability to encroach. We show that in such a structure, if both the penalties imposed for shirking are high enough and the probability of being detected are sufficient, then shirking can be prevented and the gains from cooperation potentially realized. The central question here, however, is how to divide the gains from exchange. In a repeated setting, we demonstrate that the division of rents is subject to the folk theorem: any division of rents is sustainable as an equilibrium.

This raises an additional problem: which punishment regime, and thus division of benefits between states and center, will actually be played? Can states *coordinate* on a punishment regime, which ensures maximal benefits are returned to the states? If we consider the play to be embedded in an institutional game, we show that coordinating devices, such as constitutions, can serve to minimize efficiency losses and maximize the return of rents to the states. So far, however, we have failed to address the fundamental trade-off outlined above.

In Section 4, we modify the previous model to address the fundamental tradeoff of federalism. In particular, we add three assumptions to capture the stylized design problems outlined above. First, we include the possibility that states will incur costs for exiting from a

federal structure. Second, we relax the assumption that all states are alike. Finally, and fundamentally, we assume that the ability of the central government to provide public goods is correlated with the exit costs imposed on potentially secessionist states.

This framework generates a number of interesting results. First, rents extracted by the center are increasing in states' exit costs: if it is costly for states to leave, their options are limited, as is their ability to obtain rents. Second, the center will discriminate against certain states, with some subsidizing the center more than others. Third, we consider the optimal grant of power by the states to a central authority. The answer is that states will make this trade-off by comparing the relative improvement in productivity from a stronger center to the risk of hurting bargaining power with the center. Finally, up to this point we have considered only bottom-up federalisms. Using our framework, we compare them to one in which federalism is top-down. Not surprisingly, we show that the central authority is able to keep more of the rents from a federal structure for itself under a top-down regime.

In Section 5, we illustrate our results by exploring problems from actual federalisms. We consider three cases: the problems in the United States under the Articles of Confederation, the nullification crisis during the first Jackson administration, and the problems facing modern Chinese federalism.

In Section 6, we offer some concluding remarks, discussing future extensions of this work.

2. A Model of Decentralized Cooperation

Federalisms occur when they are *both* sustainable *and* necessary. In this section we consider the conditions for the latter, while in Sections 3 and 4 we deal with the former.

Federalisms possess $N + 1$ governmental units: a center and N states. This implies there must be *some* role for each of these units. The rationale for a hierarchical structure has to begin with some opportunity for gains from hierarchy. The case of *bottom-up* federalism, in which states organize a federation, typically reflects the opportunities for *gains from exchange*. *Top-down* federalism, in which the center organizes N sub-national units, typically reflects the *gains from specialization*. But gains from hierarchy, while necessary, are not sufficient for a

federalism to arise. In particular, a number of scholars have demonstrated that *decentralized* cooperation can emerge through repeated interactions (see for example, Bendor and Mookherjee (1987), Kreps and Wilson (1990), Green and Porter (1984), Axelrod (1984)). So for a federalism to emerge, decentralized cooperation must fail.

In this section we consider when such cooperation occurs in a decentralized manner. Our model follows that of Bendor and Mookherjee in that we see cooperation as an *n-state* prisoner's dilemma. Using this model, we derive the conditions under which decentralized cooperation can occur, obviating the need in bottom-up federalism, for a central authority. We develop this model for two reasons. First, it provides a base-case from which we can later expand the model to consider centralized hierarchies. In this sense, it is a way to introduce notation and solution concepts for the later models. Second, it demonstrates the conditions under which federalism is unnecessary.

2.1. The Basic Model (Bendor and Mookherjee)

The basic model, which we call the *decentralized game (DG)*, is the infinite repetition of the following stage game. There are N players, indexed $i = 1,...,N$. In each period t a player must choose to either *contribute* or *shirk*, $A_{it} = \{C, S\}$, which is recorded by the indicator variable $k_i=1$ if the player contributes. If a player contributes he pays a cost of one, and zero otherwise. In each stage, a player gets $1/N$ of the benefits, which are simply the sum of contributions in that period modified by a parameter θ, which reflects the increase in benefits from cooperation. To make the set of strategic choices a prisoner's dilemma, we assume that $\theta>1>\frac{\theta}{N}$.

A player's payoffs in the stage game are:

$$u_{it}=\frac{\theta}{N}(\sum_{j=1}^{N} k_j)-k_i \ .$$

A player's payoff for the repeated game is simply the sum of the stage payoffs discounted by a factor $\delta\in(0,1)$ for each stage:

$$u_{i\infty}=\sum_{t=0}^{\infty} \delta^t u_{it} \ .$$

The parameter δ reflects the degree to which the players value the future. A large δ implies that players care a great deal about future payoffs, and therefore are willing to sacrifice some of today's benefits in order to garner greater benefits in the future.

A player's *strategy* describes what a player will do given all possible histories H_t of the game to that point. In particular, a player's strategy in period T depends on $A_i^{T-1}=(A_{i1},A_{i2},...,A_{i(T-1)})$ for all i. Define the set of all possible histories up to point t as $H_t=A_1^t\times A_2^t\times...\times A_N^t$. Then player I's strategy is a function $\sigma_{ik}(h_t)$ which in *each stage* maps all possible histories $h_t\in H_t$ into a *choice* $\{C,S\}$, so $\sigma_{ik}:h_t\rightarrow k_{it}=(0,1)$.

Finally, we assume that in all stages there is *complete information*. Players know the structure of the game, the history of the game to that point h_t, and the strategy being played by all other players.

2.2 Results

In infinitely repeated games, there invariably exist a multiplicity of equilibria. A number of folk theorems have demonstrated that, given sufficiently patient players, every feasible payoff set that is individually rational can be supported as a Nash equilibrium (see for example Fudenberg and Tirole (1991), Theorems 5.1 and 5.2).

For the purposes here, we are particularly interested in the conditions under which cooperation can be sustained as an equilibrium. Cooperative equilibria are defined as those in which, on the equilibrium path, all states choose C in every stage. Following the solution concept outlined above, we consider the parameter space under which cooperative equilibria can be sustained for a punishment strategy commonly referred to as *grim trigger (GT)*:

> **DEFINITION 2.1**. *A player i plays a* **grim trigger strategy** *(GT) if in each stage, he plays C if all other players have played C in every turn previously. If any other player has ever played S, then i plays S for every turn thereafter, given the opportunity.*

Under grim trigger, the players will cooperate only as long as the other player has always cooperated.

We analyze the equilibria under *GT* for two reasons. First, *GT* is suitable because it is the most extreme form of punishment that is still subgame perfect. That it is subgame perfect with complete information is straightforward: the punishment strategies are, for this game, simply Nash-reversion strategies, which means that they are subgame perfect off the equilibrium path (Morrow (1994)). In this sense, grim trigger is a *test case*, a necessary condition for cooperation to be a Nash equilibrium. If cooperation cannot be sustained under a grim trigger punishment strategy, it is unsustainable under any feasible strategy. Second, the results that follow in Propositions 1 and 2 hold for any finite period punishment phase (as shown in Bendor and Mookherjee). While analytically more convenient, *GT* yields substantively similar results to any other strategy in this class.

Using this solution method, we obtain the following characterization of sustainable decentralized cooperation.

PROPOSITION 2.1 (BENDOR AND MOOKHERJEE FOLK THEOREM). *In the DG, grim trigger can sustain cooperative outcomes if*

$$\delta \geq \frac{1-\theta/n}{\theta-\theta/n} \quad . \tag{2.1}$$

Proposition 2.1 indicates that as long as the players value the future enough, cooperation can be sustained, a result that is consistent with the many folk theorems (Friedman (1971); Fudenberg and Tirole (1991)). Further, we can solve (2.1) to define a critical group size n^*. This yields:

$$n^*(\delta,\theta)=\frac{\theta-\theta\delta}{1-\theta\delta} \quad .$$

n^* defines the largest group for which cooperation can be sustained, given δ and θ. If $n > n^*$, then cooperation cannot be supported as an equilibrium. If the benefits of cooperation or the value the players place on the future declines, cooperation will only be sustainable in smaller and smaller groups. More importantly, however, Proposition 2.1 allows us to explore the relationship between group size, benefit-cost ratios and cooperation.

PROPOSITION 2.2 *Cooperation is more difficult to sustain as n gets large, δ gets small, and the benefit to cost ratio declines.*

The intuition behind this set of results is fairly strong. Take first the discount factor. For cooperation to be sustainable through repetition, the threat of lost future benefits must outweigh the temptation to shirk. If states do not sufficiently value the future, *ceteris paribus*, then this threat loses its force. In the extreme, if a state only cares about its utility in the *current period*, then its choice is that of the one-shot game, and it will always defect. The intuition behind the benefit to cost ratio is similar. The threat of lost future streams of net benefits is much more potent as those net benefits increase. The logic behind n^* is more subtle. When a state decides not to contribute, then it gets $\frac{n-1}{n}\theta$ immediately. The fact that it does not contribute is reflected in the numerator $(n - 1)$. As n grows, the $n - 1$ tends towards n, and the fact that it did not contribute becomes insignificant; the benefit from defecting grows. Thus, the incentive to defect is larger and larger as the number of states increases.

Combined, Propositions 2.1 and 2.2 give us necessary but not sufficient conditions for the emergence of sustainable federalism. They suggest that federalism will only occur if the gains from cooperation cannot be captured in a decentralized structure. Often times, a decentralized punishment regime will be sufficient to capture these gains. However, as the number of states grows large, the net benefits grow small or the states do not sufficiently value the future, then a centralized solution will be necessary. Unlike in Sections 3 and 4, where we show the conditions under which a necessary federalism is yet *unsustainable*, here we show when a federalism is *unnecessary*.

3. Centralized Provision of Public Goods

In the previous section, we saw that decentralized cooperation requires a small number of states, high discount factors and large gains from trade. If there is incomplete or imperfect information, cooperation becomes even more difficult to sustain (Green and Porter (1994);

Bendor and Mookherjee (1987)). If these conditions are not met, a central monitor can often allow states to capture the potential economic benefits of cooperation or centralization.[2]

Introducing a central monitor has the potential to obviate the shirking problem by the states. If the center is granted enough power to police, then the threat of penalties might fill the role of decentralized punishments in the *DG*. At the same time, however, creating a central authority, as in any hierarchical structure, means that there is potential that some of the rents from hierarchy will be extracted by the center. In this section, we modify the *DG* to take account of cases in which a central government is required. Before introducing all of the stylized facts discussed in the introduction, we consider a basic model in which the center and the states choose strategies of contributions and goods provision in a world in which there is *imperfect observability* of state contributions.

3.1 A Basic Model of Centralized Government

Here we offer a model of centralized government we call the *centralization game (CG)*. The game is the infinite repetition of the following stage game. Unlike the *DG*, the *CG* has $N + 1$ players, n states indexed by $i = 1,...,N$, and a *central government* called *C*. In each stage, the sequence of moves is follows. Each state first picks a strategy from the action set $A_i = \{C, S, E\}$. If a state chooses *C*, it chooses to *contribute*, which means it provides a numeraire resource indicated by the variable $k_{it}=1$. If the state chooses *S*, it *shirks* by contributing nothing, so that $k_{it}=0$. If the state chooses *E*, it *exits* the federalism. If the state plays *E* in period *t*, then $k_{it}=0, k_{is}=0$ for $s>t$, and $s_{it}=0$, where s_{it} is an indicator variable indicating it has exited in that period.

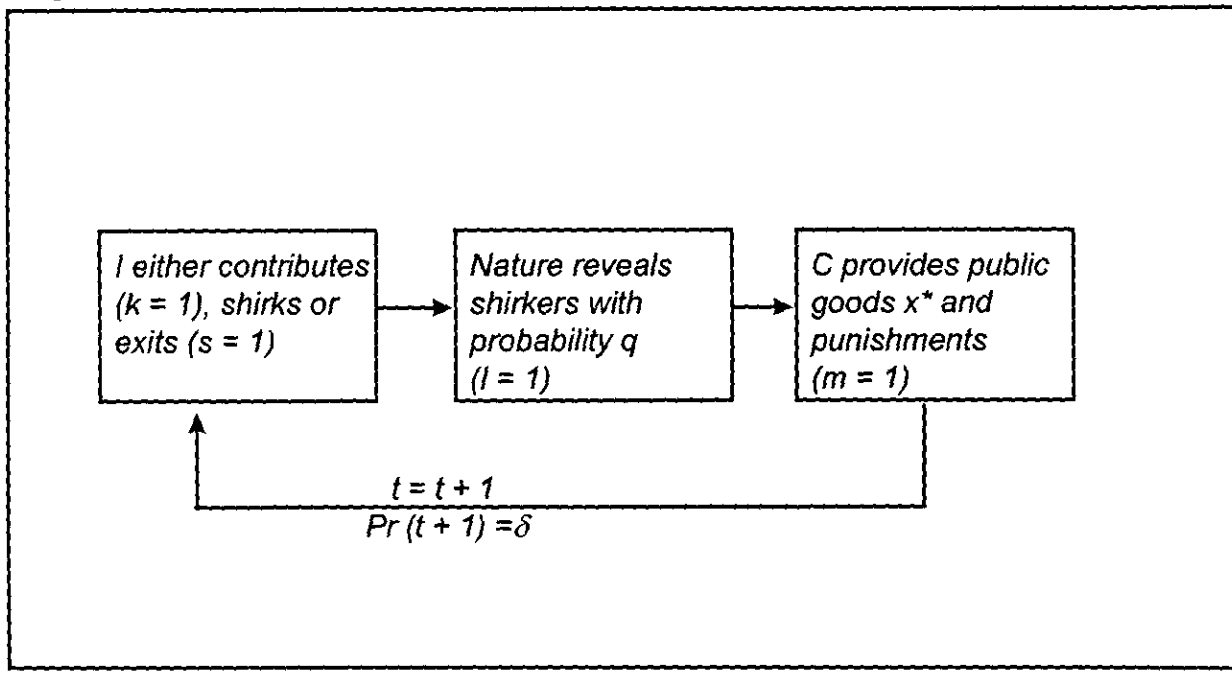
Once the states have decided whether to contribute, shirk or exit, a non-strategic player reveals shirkers with probability *q*. If a state is revealed to be shirking, then the indicator variable l_{it} is set to 1; it is 0 otherwise.

---

[2] Note that here, the *benefits of centralization* could be more than simply the gains from exchange; often a central government is actually more efficient in carrying out certain tasks. In terms of the model, this would means that $\theta_C > \theta_{DG}$. In eitherBoth cases, — if federalisms occur for the purposes of efficiency and scale, and thoseor to capture the benefits of trade, — still raises the same design issues studied here.: states still have incentives to under-contribute and the central government still has an opportunity to extract rents, so we collapse these two cases into one.

The third move in the game is made by the central government *C*. *C* has two actions in this stage. First, it must choose some payment *x*, the amount returned to each state. As in the *DG*, these payments are modified by a production or public goods technology which is parameterized by θ. Second, *C* must choose a strategy for meting out fines *f*. This choice is represented by the vector **m**, where the typical element of **m** m_{it} is 1 if a fine will be levied against state *i* and 0 otherwise. Finally, the payoffs are determined and the stage ends. The full extensive form of the *CG* is pictured in Figure 3.1.

Figure 3.1 Sequence of Play in CG Stage Game



Payoffs for each state in the stage game are the net of the payment made by *C* modified by θ and its contribution and fines:

$$u_{it}=\theta x-k_i-fm_i \ .$$

C's payoff is simply the net of the sum of contributions and fines less the payments made to the states:

$$u_{Ct}=\sum_{i=1}^{N}(k_i+fm_i)-nx \quad .$$

As in the *DG*, the repeated payoffs are the stage payoffs summed over all the periods discounted by a factor $\delta\in(0,1)$, where the parameter δ has the familiar interpretation .

Strategies in this game map histories into the choices in period *t*. Let $L_{it}=(0,1)$ indicating whether a player has been revealed to be shirking in period *t*, $M_{it}=(0,1)$ indicating whether a player has been fined in period *t*, and $h_t=L_{1t}\times L_{2t}\times\ldots\times L_{Nt}\times M_{1t}\times\ldots\times M_{Nt}\times x_t$ for all *i* still playing. Then define the set of all possible histories to period *t* as $H_t=h_0\times h_1\times\ldots\times h_t$. Then a state strategy $\sigma_{it}(H_{t-1})$ is a map $\sigma_{it}:H_{t-1}\rightarrow(C,S,E)$. Similarly, a central government strategy $\sigma_{Ct}(H_{t-1},,\mathbf{l}_t)$ is a map $\sigma_{Ct}:(H_{t-1},\mathbf{l}_t)\rightarrow(\mathbf{x},\mathbf{m})$.

3.2 Results

To solve the *CG*, we once again employ *subgame perfection*, so that players are playing optimally for every subgame. The set of equilibria is given by:

PROPOSITION 3.1 (FOLK THEOREM). *Suppose* $\theta>\frac{1}{\delta}$ *and* $qf>1$. *Then* $x^*\in[\frac{1}{\theta},\delta]$ *can be supported in equilibrium if in every period the following be the strategies played by the players:*

> *(i) states: contribute in every period if C has provided x* in all previous periods; exit otherwise. Shirk if C ever fines a player not revealed to be shirking.*
> *(ii) center: pay benefits of x* to each state. Fine all states that are caught.*

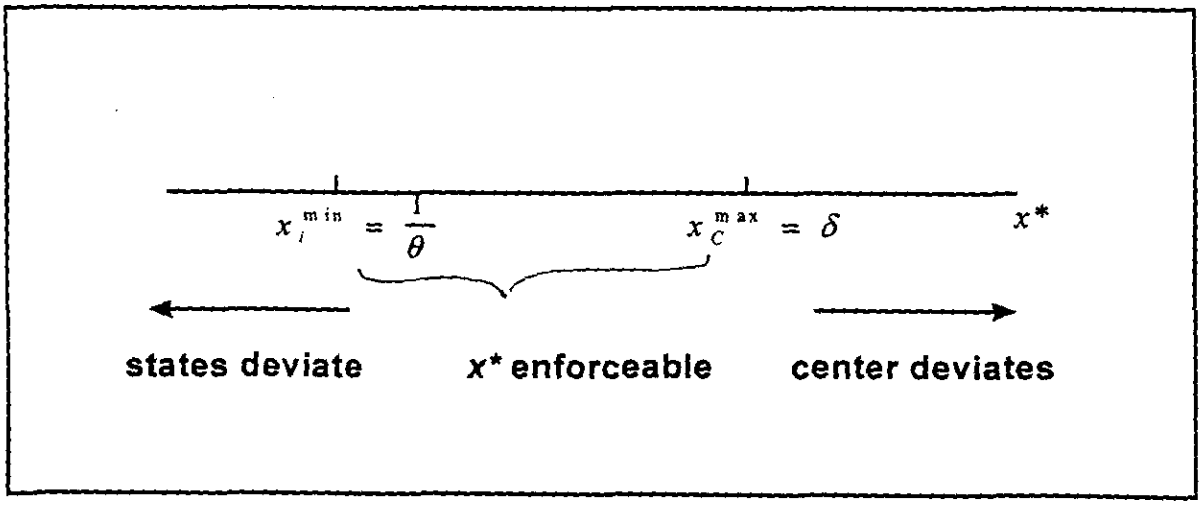
Proposition 3.1 has a number of implications for federalism. First, the second condition, $qf > 1$ means there must be a sufficiently high probability of shirking being detected *and* sufficiently high penalties available. If this is the case, then all states will contribute. The rationale is that in equilibrium, if a state shirks it will get the normal payment, but will be fined *f* with probability *q*. Thus, a state will shirk if the expected fine for shirking is less than its costs of contributing, which is one. A central authority imbued with enough power can therefore *prevent shirking*. Notice that this is a knife-edge result: because the parameters *f* and *q* are exogenous, either all states will shirk or none will. This assumption thus defines a necessary condition for a

stable federalism: the center must be given a strong enough hand to detect and punish potential shirkers.

Second, the condition $\theta > \frac{1}{\delta}$ implies that as before, *there must be sufficient gains from exchange to motivate a stable federalism.* The logic, however, is different than in the *DG*, in which the benefit stream alone had to prevent individual states from shirking. In this case, the benefits have to be analogous, but they have to be high enough to prevent the center from appropriating all contributions. Consider the center's calculus: in equilibrium, the center collects contributions from all of the states. Its choice is between taking all the contributions in the current period for itself, and losing all future payments, or continuing to receive an ongoing payment from each state. The condition states that for the center to be sufficiently motivated, it must pay out at most δ (in other words, must receive a minimum of $1 - \delta$) in each period, or it will appropriate all of the contributions for itself, causing a breakdown in the federal structure.

Third, the *CG* raises an important design problem for federalism: *the division of benefits between the center and the states*. The range of possible x^*'s has the following rationale, pictured in Figure 3.2. As we have just observed, if the states try to enforce payments to themselves of greater than δ, the center will choose instead to appropriate all the contributions. Similarly, if the center tries to make too low a payment to the states (less than $\frac{1}{\theta}$), the states will choose to exit:

Figure 3.2 Equilibria in the *CG*



they will be paying out one in every period and receiving less than that in return. Any range in between these two bounds $[\frac{1}{\theta},\delta]$ is enforceable, by the familiar punishment codes of repeated folk theorems (Abreu (1982)). In this region, both the center and the states are weakly better off than under no federalism, and therefore can be induced to conform to x^* by a set of equilibrium punishments.

This introduces a new issue for studying federalism: how are the rents from cooperation allocated? In this section, there is little that can be said about such allocations.

Finally, in terms of total social welfare, *all allocations are not equal.* Let us use a fairly loose definition of social welfare as being the sum of benefits to all parties. Then we can calculate the social welfare in the following way. In equilibrium, the states get $\theta x^* - 1$ and the center gets $n(1 - x^*)$ in every period. Thus, the per-period total welfare is $nx^*(\theta - 1)$. This term is strictly positive in equilibrium since $\theta > 1$. However, it is obvious that the social welfare is *increasing in* x^*. The reason is that the production technology benefit only accrues if C supplies public goods. Thus, every unit for which the center collects revenue but does not supply goods to the states represents an opportunity cost in terms of public benefits forgone. Thus, any allocation in which $x^* < \delta$ represents a dead-weight loss to society.

3.2 Centralization and Coordination

As noted above, if the states do not have a coordination device, then it is impossible for the analyst to say which of the multiplicity of equilibria will arise in the *CG*. Equilibria in which the states force the center to take minimal rents and equilibria in which the center appropriates all of the rents—resulting in no improvement in social welfare—are equally tenable. For bottom-up federalism, states' inability to *coordinate* on a punishment strategy mean that the division of rents is indeterminate. Institutions, however, potentially provides a way out of this quandary. In bottom-up federalism, the states get a say in the *design of the institutions of federalism*. This means that they are afforded an opportunity to predetermine the punishment strategy upon formation of the federal structure. When erected prior to playing the federalism game, a constitution can serve as a focal, coordinating device by determining precisely what constitutes central encroachments.

To explore this possibility, we consider a modification to the *C*, by introducing the following *pre-play* stage. Prior to the initiation of the repeated *CG*, the states must choose by majority rule a *punishment cutpoint* x^C. This cutpoint is the trigger point for state punishments: if the center ever provides less than x^C, the states will initiate the punishment of the center.

To find the subgame perfect equilibrium choice of x^C, we use backwards induction, with the repeated *CG* equilibria stated in Proposition 3.1 as the input to the pre-play stage. Proposition 3.2 elaborates the equilibrium choice of the states.

PROPOSITION 3.2. *Suppose there exists an equilibrium to the CG. Then the states will choose an equilibrium cutpoint* $x^* = \delta$.

Proposition 3.2 demonstrates that, if given a chance to coordinate, the states will choose to play the equilibrium that maximizes their payoffs. They will choose a punishment strategy that at once ensures that the center will have the incentive to participate, but just barely, and will capture the remainder of the rents for themselves. In other words, the opportunity for establishing focal strategies gives an institutional advantage to the states. This is precisely the role that can be played by a clear delimitation of federal authority and responsibility and states' rights in a constitution (Weingast (1997b)). Further, using the results above, we also know that the pre-play equilibrium results in the greatest provision of public goods and therefore the highest amount of social welfare possible in a centralized system. Thus, the constitutions not only shift the institutional balance, but improve social welfare.

4. Model of Federalism with Endogenous Institutional Choice

In this section, we expand the model to take account of some of the stylized facts and puzzles we presented in Section 1. In the previous models, we demonstrated two facts. First, if gains from trade exist, decentralized cooperation among sub-national units[3] may be possible. But as the model in Section 2 shows -- as do a number of more general results from the game

[3] For the purposes of this exposition, we use the terms "sub-national unit", "state" and "sub-unit" interchangeably.

theory literature (e.g. Bendor and Mookherjee (1987); Fudenberg and Tirole (1991)) — a decentralized solution can only obtain if the number of states is small, the units value the future enough, information is relatively good, and/or the relative benefits from trade far outweigh the costs. If these conditions do not hold, then in general, a decentralized system will not yield the benefits of cooperation; the incentive to defect outweighs the potential benefits of cooperation and thus the threat of punishments for intransigence.

In section 3, we generalized the model, providing for a central monitoring agent called the "center." The center polices the sub-national units and is endowed with an institutional basis for exacting fines or punishments for shirkers. In this case, we derive a folk theorem which shows that for sufficiently high fines and a sufficiently large probability of detection, states will uniformly contribute. Further, ignoring any pre-play coordination by the states, we also show that any division of the gains from trade between the states and the center can be sustained in equilibrium. If we include such a pre-play stage, the result will be the equilibrium in which the states extract all of the extraordinary gains and the center is left at its indifference point between defecting from the public goods provision regime and providing those goods.

In this section, we generalize the model through three modifications. First, we introduce exit costs for the sub-national units; once they have been induced to participate in the pre-play, institutional design stage, sub-national units can only exit from the system at a cost. In practice, these costs can be thought of as representing any real economic costs (as opposed to opportunity costs) that states incur when they leave. The primary example of such costs are the expected use of force upon secession. Second, we modify our original conception of a strict public good. In this case, although there are some positive externalities to providing services at the national level, we allow the benefits to be differentially spread across the units. Both of these modifications mean that we introduce a degree of heterogeneity in preferences: states can be of an infinite continuum of cost types. Finally, we embed the fundamental trade-off in the model. As we noted earlier, the problem of institutional design is that the states want to provide resources for the central unit to provide public goods. In so doing, however, this also increases the power of the center to manipulate the states and extract rents. Thus, we add in the institutional design stage a grant of institutional power. This power has two implications. On

the one hand, it increases the ability of the center to provide public goods; on the other hand, it increases the exit costs for all states, although it maintains the order of such costs. Again, we can turn to the case of providing for defense. In this case, the more institutional power the states grant the center—for example, the ability to call up state militias and deploy them for national defense—increases the value of such defense for a given level of resources. At the same time, however, it also improves the ability of the center to punish secessionists, making it harder for such recalcitrants to leave.

We consider a number of variants of this model to explore alternative institutional designs. Broadly speaking, we reach a number of theoretical conclusions. First, the introduction of exit costs shifts some of the rents to the center. Second, in designing a bottom-up system, states face a fundamental trade-off. On the one hand, they want to increase the productivity of federal resources spent for public goods. On the other hand, increasing productivity also increases the exit costs, meaning that the center will capture a larger proportion of the benefits. In facing these challenges, then, we show that the states will design a system that makes this trade-off by equating marginal benefits from greater productivity with the marginal increase in exit costs. Finally, if the system is designed by the center, exit costs will be higher, the proportion of rents extracted by the center will increase and there will be larger deadweight losses.

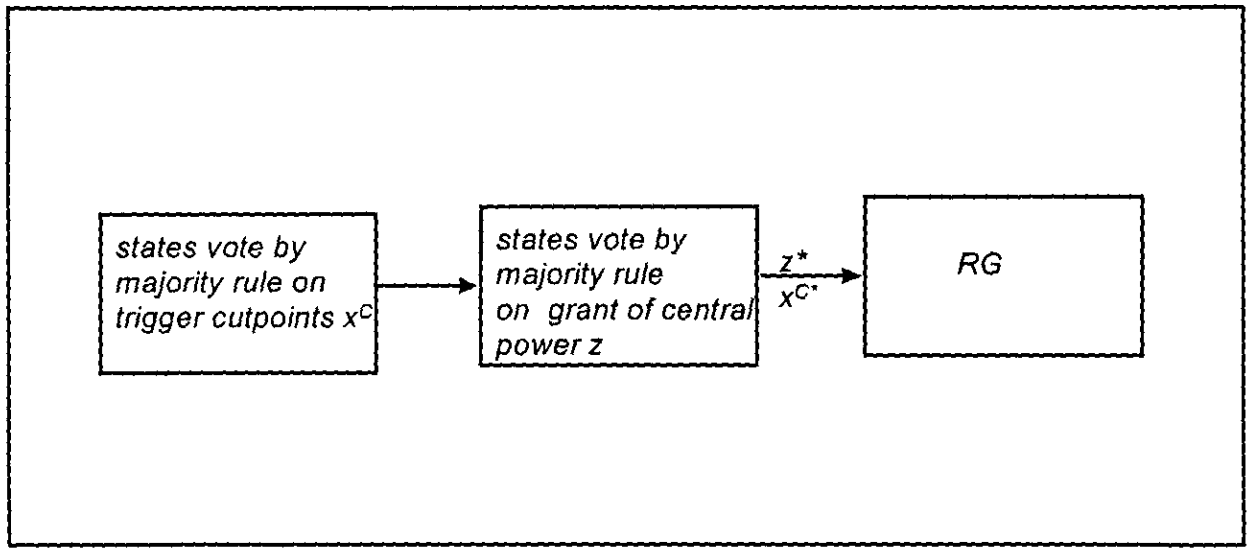
4.1 Model

As previously, we assume that there are $N + 1$ players, a center which we label C, and N states indexed by $i = 1,\ldots,N$. The game has two stages. In the first "pre-play" stage, which we call the *institutional game (IG)*, where $t = 0$, the states confer to choose an institutional design. States make two choices. First, as before, they pick a *punishment strategy cutpoint* x_i^C. Note that since there is heterogeneity among the states, these cutpoints might now differ. Since the only source of heterogeneity that we assume is in the exit cost functions $c_i(z)$, however, all differences among the states' cutpoints will be functions of those potentially unique exit costs. As before, we envision this choice as the embodiment of rights and responsibilities in a constitutional document which gives the sub-national units an opportunity to coordinate on

a punishment strategy. Second, by *majority rule* the states choose a parameter z, which is an argument in both the exit cost functions $c_i(z)$ and the center's production transformation function $\theta(z)$. Further, at this stage, any state can choose not to participate, if the choices make the sub-national unit worse off than under no cooperative agreement. Notably, this structure, in which states can both choose to participate and set the institutional standards by some preference aggregation rule in which no single player is decisive, implicitly means we are modeling *bottom-up federalism*; the states are designing rules to sustain cooperation.[4]

The next game is the *repeated game (RG)*. As in the previous sections, this game is a repeated moral hazard model. The sequence of moves is the same as in Figure 3.1, with some

Figure 4.1 Sequence of Play in IG



modification in the payoffs and parameters. First, the states choose one of three actions $A = \{C, S, E\}$. If a state chooses C, it means the state *contributes* one unit to the center. If a state chooses S, the state chooses to *shirk* and contributes zero. If a state chooses E, it also contributes nothing and chooses to *exit* or secede from the federal system. The indicator variable k_i=1 if a state contributes and 0 if it does not. Further, a state's choice of exiting or not is designated by the indicator variables s_i which equals 1 if the state chooses to exit and zero

[4] We assume an open-ended agenda process here. Specifically, we assume that any element in the core is a possible equilibrium outcome.

otherwise. If a state chooses to secede, then it also incurs a cost, which is a function of the center's institutional power granted in the *IG*, $c_i(z)$. We assume that $c_i(z)$ is an increasing, convex function in z. We also assume that the $c_i(z)$'s are ordered in z. In other words, if for any z $c_i(z)>c_j(z)$, then $c_i(z)>c_j(z)\forall z$. This simply means that states costs *relative* to each other are the same. Finally, exiting means that the state no longer participates in the game, incurring no costs or benefits in later stages. The second step in the stage game is that a non-strategic player reveals shirkers with probability q which is determined exogenously. Those players revealed to be shirking are indicated by a value of 1 of the indicator variable l_i. All players observe only the vector $\mathbf{l} = (l_1,l_2,...,l_N)$, so potentially, some shirkers go undetected by the center and sub-units. The third move in the game is made by C, the central government. C chooses a *payment vector* $\mathbf{x} = (x_1,x_2,...,x_N)$, which is the amount of payments made to each sub-unit. As before, the payments to the sub-units are modified by a production transformation technology $\theta(z)$. Note that in contrast to the model in Section 3, this production technology is a function of the institutional grant of power z made in the *IG*. We assume that $\theta(z)$ is an increasing, concave function in z, so that there are diminishing returns to power. C also chooses a *shirking penalty strategy* $\mathbf{m} = (m_1,m_2,...,m_N)$, which is a vector of indicators indicating if an exogenously determined penalty or fine f will be levied against each sub-unit i. Finally, payoffs for the stage are determined and the stage ends.

The payoffs of the actors are as follows. The utility function for state i is:

$$u_i=\theta(z)x_i-k_i-fm_i-s_ic_i(z)$$

The amount returned to a state x_i by the center is modified by the production parameter $\theta(z)$. The state then makes a contribution designated by k_i. The state also pays a fine equal to f if $m_i = 1$. Finally, if $s_i = 1$, if the state chose to exit, then it pays an exit cost $c_i(z)$. The center has a utility function given by:

$$u_C=\sum_i k_i-x_i+fm_i+\alpha s_ic_i(z)$$

Thus, the center gets the sum of contributions less its outgoing payments to each state. Note that we also include benefits to the center from exiting. The reason these benefits are included is that when the state enters a federal bargain, and carries with it exit costs, its bargaining power upon exit is reduced. Therefore, to exit, the state must give up something; it must pay the center to exit. The center might not get all of the benefits, which is reflected in the parameter α. An instructive example would be a secession followed by a threat of war. With complete information, an exit of this type would only take place if the center could be discouraged from entering a war. Thus, the state would have to make a payment in order to prevent such an outcome. For now, we assume $\alpha = 1$, an assumption which we will relax later.[5]

The repeated payoffs are simply the stage payoffs summed over all the periods that the player is playing discounted by a factor δ. Thus, the repeated payoffs are:

$$u_{j\infty} = \sum_{t=0}^{\infty} \delta^t u_{jt} \qquad j = C, 1, \ldots, N.$$

We assume that players choose actions that maximize the expected value of $u_{j\infty}$.

Strategies in these games describe completely contingent plans. First, in the *IG*, states must choose a vote for all possible z's and a cutpoint x^C. So in the stage game a strategy is a map $\sigma_{iIG}(g_1,g_2;\theta(z),c_i(z))$ where the functions are defined as $g_1:\phi_1 \in \mathbb{R}^+ \rightarrow (1,0)$, a map from the set of possible z's into *votes* either "yes" (1) or "no" (0), and $g_1:\phi_2 \in \mathbb{R}^+ \rightarrow (1,0)$, a map from proposed cutpoints into *votes* also either "yes" or "no." For the states in the repeated game, a strategy $\sigma_{iRG}(H_\tau,c_i(z))$ maps all possible histories of moves, including those of the *IG* into the action space *{C, S, E}*. Thus, for a given period T, let $h_t = \phi_1 \in \mathbb{R}^+ \times \phi_2 \in \mathbb{R}^+ \times (C,S,E) \times x \in \mathbb{R}^+$ and $H_T = h_1 \times h_2 \times \ldots \times h_T$. Then $\sigma_{iIG}: H_{T-1} \rightarrow (C,S,E) \ \forall T$. Similarly, the center C's strategy maps histories into choices of **x** and **m**, given **l**, $\sigma_{Ct}:(H_{t-1}, \mathbf{l}_t) \rightarrow (\mathbf{x},\mathbf{m})$.

---

[5] Another way to see this point is to imagine a toy game that is initiated upon exit in which the state credibly offers either c or zero to the center for exit, and then the center can either impose the costs or not. In this situation, the state will pay the center exactly its exit costs, so a transfer is made.

4. 2 Results

To solve this game, we use the equilibrium concept of *subgame perfection.* In this context, that means players are playing optimal strategies at each point for every point forward. In implementing subgame perfection, we use backward induction, solving first the *RG* and then, conditional on the results from that solution, we solve the *IG.* Notably, *within* the *RG,* we cannot use backward induction, since the game has a positive probability of continuing at every point. Instead, we try to characterize classes of equilibria by positing the equilibrium strategies of the players and testing whether those strategies are optimal, given the other players' strategies.

Equilibrium of the RG. In Proposition 1, we then characterize an equilibrium for the *RG.*

PROPOSITION 4.1A. *Suppose* $fq>c_i(z)>\frac{1}{1-\delta}\forall i,z$, *and* $\theta(z)>1\ \forall z$, *then* $x_i^* \in (\frac{1-c_i(z)(1-\delta)}{\theta(z)}, 1-(1+\delta c_i(z))(1-\delta))$ *can be supported as an equilibrium with the following strategies:*

(i) Center: $m_i^*=l_i$, $x_i^*=x^{C*}\ \forall i$ *if* $x_i^{C*}<1-(1+\delta c_i(z))(1-\delta)$.

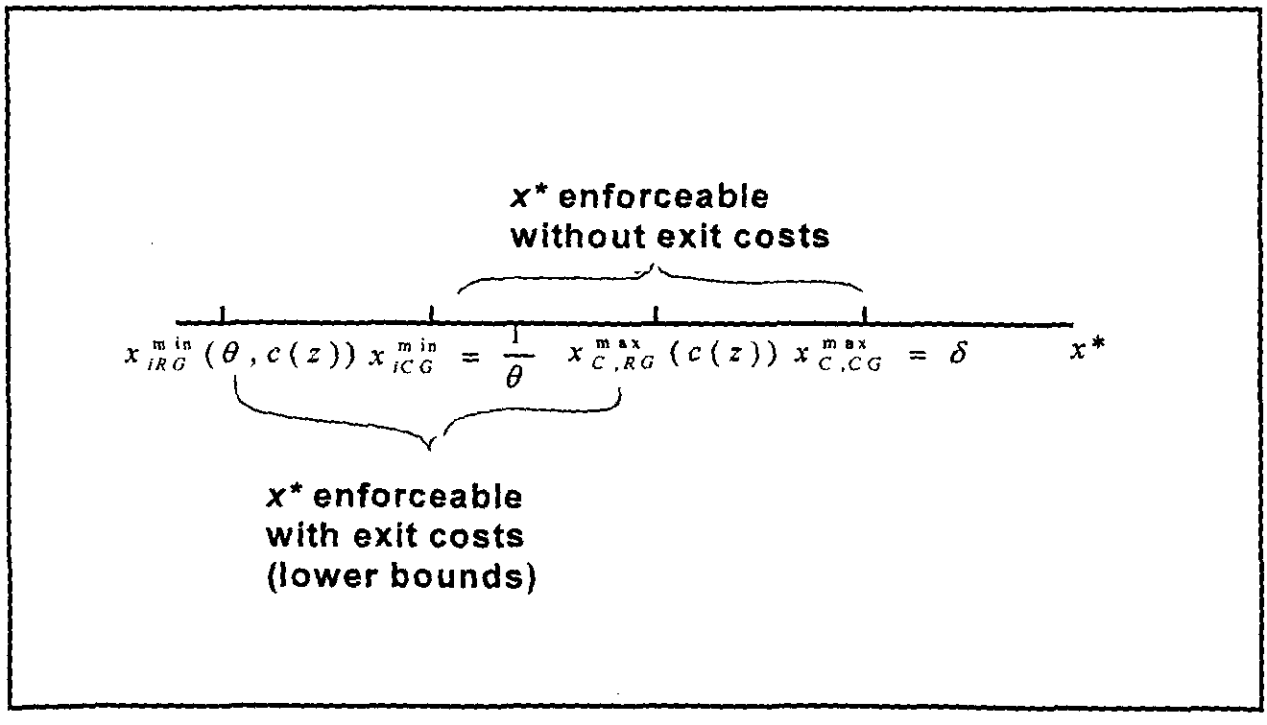
(ii) States: play C in every period except if center ever provides $x_i<x^{C*}\geq\frac{1-c_i(z)(1-\delta)}{\theta(z)}$, *then play E. Play S in all periods if* $l_i\neq m_i$.

Proposition 4.1a gives us insight into the ongoing dynamic that occurs between the center and the states. First consider the condition given $fq>c_i>\frac{1}{1-\delta}\forall i$. What this condition means is that the expected fine for shirking is large. Of course, that could occur if *either f,* the fine, or *q,* the probability of getting caught when shirking, is sufficiently large. As in the earlier model, if the expected fine is sufficiently large, the center can prevent shirking by all sub-national units. If not, however, the federal system will break down since all states will have an incentive to deviate. Further, the condition $\theta(z) > 1$ for all z, means that there must be some gains from exchange between the sub-units.

Second, if the costs of exiting are sufficiently high, the states will also have an incentive to not exit, although the center does not pass on all of the rents to the sub-units. This is an important distinction from the earlier model, for it indicates that *exit costs* can shift economic and institutional power from the states to the center as shown in Figure 4.2. To see this note two points. First, in comparison to the *CG*, the, both the upper and lower bounds, on x_i^* are lower. Second, both of the cutpoints are decreasing in $c_i(z)$. What will cause states to exit? Only if the center starts extracting *more rents* than the exit costs will the states exit. In other words, as long as the center continues to pay out the excess benefits to the states, the states will not have an incentive to leave the federation. Thus, if the benefits are sufficiently large in relation to the exit costs, a stable federalism can be sustained. An interesting aside is considering what happens if we relax the assumption that $\alpha \doteq 1$. If α i small, when the center does not incur any benefits from exit, say for example, in a stale-mated civil war, then the game for the center is reduced to the *CG*; the upper bound will once again be δ.

In addition, as in the *CG* without pre-play, the *RG* suffers from multiple equilibria. As before, there is a range of possible equilibrium outcomes that can be supported as divisions of the rents. Once again, it is not possible to say which f these equilibria, and therefore will be played in

Figure 4.2 Equilibria in the *CG*



practice.

Further, the *heterogeneity* in the states' cost functions means that the center can *price discriminate*. For those states that have a large cost of exiting, the center will have to pay less to induce them to continue in the federation.

Finally, consider the shirking punishment strategies. The center gets utility from fines, so it has an incentive to fine all states, whether shirking or not, in order to increase its utility. In the one-shot game, the center will fine all states. But the states can condition this incentive in repeated play. In particular, the more contributing states that the center fines, the more all contributing states will have an incentive to deviate (shirk). Thus, the states can credibly commit to punishing any deviation from a direct mapping of caught shirkers to fines, and the center, given the benefits from ongoing execution of the equilibrium shirking penalty strategy along with continued rents, will not have an incentive to extract such "inappropriate fines."

Equilibrium of the IG. Given these results for the *RG*, we can now turn to the more interesting question of how the states will design institutions if given the opportunity. Here, the states must pick a cutpoint x^C to trigger punishments, which as we have seen previously, must be less than $1-(1-\delta)(1+\delta c_i(z))$ in order to motivate the center to not take the payments and live with a mass exit. The states must also choose z. As we have mentioned before, this choice is one of the fundamental trade-offs in federalism. On the one hand, assuming that they can motivate the center to return a significant part of the payments to the states, then a higher z will mean a higher θ, yielding larger benefits per unit for the states. On the other hand, choosing a higher z also increases the exit costs for every state, meaning the center can extract more of the rents from the states, as we saw above. Proposition 4.1b gives the formal statement of the states' choices in this context.

PROPOSITION 4.1B. *Suppose the conditions from Proposition 4.1a hold, the players are playing the equilibrium to the RG stated in that proposition, and* $\theta(z_M)(1-(1-\delta)(1+\delta c_1(z_M)))-1\geq 0$, where $c_1(z)>c_i(z)\forall i>1, c_2(z)>c_i(z)\forall i>2,\ldots$ *and* z_M *is the ideal z of the median voter. Then the following constitutes a unique sub-game perfect equilibrium to the IG:*
(i) $x_i^{C*}=1-(1-\delta)(1+\delta c_i(z_M))$
(ii) z^* *is chosen to satisfy condition (*):*

$$\frac{\theta'(z)}{\theta(z)} = \frac{(1-\delta)c_i'(z)}{(1+c_i(z)(1-\delta))} \qquad (*)$$

where the subscript M denotes the median state.

Proposition 4.1b raises a number of interesting points about the design of federal institutional structures. First, as in the previous model in Section 4, here the states are in agreement on a punishment strategy: *all states have an interest in limiting central rent extraction* so they "offer" the minimal credible rents to the center and punish any deviations from such an offer. As we argued earlier, in this expanded model, the drafting of institutional rules such as constitutions which clearly delimit jurisdictions, responsibilities and authorities, can act as a guide so that states can coordinate on a *trigger* point in which all states' benefit.

Second, as noted earlier, *the center extracts greater rents when there are exit costs and these rents are increasing in those costs*. There is nothing in the institutional design stage that can be done about the relatively greater resources extracted by the center. Because the threat to secede is not credible if the exit costs are high enough, the center can always extract some of the resources granted to it, assuming that $\theta(z)$ is high enough to guarantee state participation. Further, we can characterize the rents extracted by the center as *opportunity deadweight losses* from the "first-best" if the center returned all resources to the states. Every unit of resources which is *not* returned to the states fails to garner the benefit of the transformation technology $\theta(z)$. We can specify the exact losses as $n\theta(z_m)(1-\delta)(1+\delta\bar{c}(z_M)))$, where $\bar{c}$ is a statistic that averages the exit costs evaluated at z_M, which is increasing in the exit costs.

Third, *states will not participate unless the benefits are high enough ex ante*. The condition $\theta(z_M)(1-(1-\delta)(1+\delta c_1(z_M)))-1 \geq 0$ in Proposition 4.1b is a condition which guarantees participation by all the potential states. The intuition behind this condition is that, at a minimum, the highest cost state must get a weakly positive benefit in each round (and thus over all the game) in order to participate. Else, it will not be part of the bargain. This, then, defines an additional condition for a stable federalism. Although we have forced this assumption on the model for expositional simplicity, an alternative interpretation is possible. Suppose, instead that we do not require universal participation, so that there are some states which do not satisfy the above assumption. In this case, the pre-play stage, in which states can opt out, acts as a *screening*

device, so that only those states of *high enough type* (in other words, low enough exit costs) will enter. Thus, the design stage acts as a selection mechanism to centralize the distribution of admitted states. To make an even further unmodelled conjecture, we can imagine a model in which the participation stage is ongoing. Suppose further that over time, exit cost functions of excluded countries undergo a series of random shocks. Then, with an ongoing participation stage, even fixing the one-shot development of endogenous entry and institutional rules, then we would expect that over time, a number of the originally excluded states will eventually gain admittance to the federation, particularly since these will be the low-rent, high contribution members in all likelihood.

Fourth, *states are discriminated against differentially*. In this model, the center extracts some benefits and therefore is willing to maintain its role in the hierarchical structure. Further, it is willing to do this as long as the *sum* of the benefits from maintaining the federal bargain are greater than a one-time departure. But the way the states contribute to this incentive pool differs by state. In particular, those states that have a high exit cost, contribute relatively more to the centers' subsidy. This means that low exit cost states can enjoy almost the full benefits from central goods supplied, while the high cost states might enjoy almost none. Of course, the participation constraint puts a lower bound on how small these extractions can be; indeed, the states will not participate if they will be worse off *ex ante*. But the relative welfare of these states can be much lower than some of the lower cost states. Further, notice that as the *spread* of cost functions converges to the degenerate case in which $c_i(z)=c_j(z) \ \forall i,j$, the discrimination goes to zero. Thinking of exit costs being inversely related to the size of a state is a useful potential illustrative manifestation of this phenomenon. Suppose there is a continuum of states in terms of their size, which defines a set of cost functions which are decreasing in size. Then the result here states that smaller states (say Montana and Rhode Island) will contribute disproportionately to the central subsidy than larger states (say California and New York).[6]

Fifth, *the choice of z is the optimal choice for the median voter*. The logic of this result is as follows. Each state has a z which is optimal based on its unique maximization problem. It's

[6] Of course this is conditional on the fact that all contribute equally which is perhaps an oversimplification which would temper this result.

uniqueness is based on the uniqueness of its exit cost function. As we discuss below, states with high cost functions will want to pick a lower z than states with low cost functions, since their marginal benefits are the same, but the marginal costs of high cost types will (weakly) expand more rapidly. The solutions to these individual maximization problems, however, means that we can characterize the states on a single spatial dimension with respect to their optimal z's, which can be considered induced *ideal points*. Assuming the participation constraint is satisfied for all states, then, with majority rule we are led to a median voter result by Black's Median Voter Theorem (MVT).

Sixth, *there are welfare losses incurred through aggregation*. Here we make the comparison against a case in which not only are the goods provided to each state x_i allowed to vary but also the institutional strength with respect to each state, say z_i. In this case, the problem in the *IG* amounts to providing a single public good via the center for each state, where the center is bound differentially by rules for each individual state. The efficiency loss, then, from imposing a *single set of institutional rules*, embodied in a single z^*, amounts to the loss from imposing the constraint that $z_i=z_j \ \forall i,j$. Figure 4.3 illustrates this point. The continuum z arrays all of the states according to their induced ideal points z_i. If all of the states are able to determine the center's strength with respect to their own state, then the z^* curve will be an increasing function of z. A single institutional rule imposes the condition that z^* must be flat for all i. The loss, then, is the area above the curve for $z_i<z_M$ and below the curve for $z_i>z_M$, which is strictly positive if $z_i \neq z_M$ for some $i \neq j$.

[FIGURE 4.3 about here]

Finally, condition (*) defines more explicitly the trade off between increasing the exit costs and improving productivity. The familiar intuitive interpretation for this condition is that the median will set z so that the *marginal benefit of increasing productivity will be equated to the marginal reduction in benefits from increasing exit costs*. These factors are normalized by the basic level of costs and benefits, $(1-c_i(z))$ and $\theta(z)$, respectively. Since $c_i(z)$ has increasing marginal costs to exiting which reduces the rents extracted by the sub-unit, any deviation from z^* will lead to a smaller total benefit. This means that if z is too low, then the federalism will under-

produce although a greater *proportion* of the rents will go to the states. If z is too high, there will be more efficient production; however, a small proportion of the contributions will go to providing public goods, meaning a lower welfare and exit by high cost states.

To see this more clearly, an example is instructive.

Example 4.1 (preliminary). Let $n = 3$, $\theta(z)=z^{\frac{1}{2}}$, $c_i(z)=\frac{1}{10+5i}z^2$, $\delta = 0.5$.
Here we can solve for condition (*) for the median, $i = 2$. First, writing down the first derivatives of the functions we get $\theta'(z)=\frac{1}{2z^{\frac{1}{2}}}$ and $c_2'(z)=\frac{1}{10}z$. So the condition can be written:

$$\frac{1}{2z}=\frac{0.5\frac{1}{10}z}{1+0.5\frac{1}{20}z^2} \Rightarrow z^*=\frac{2\sqrt{30}}{3}.$$

Now we can check the participation constraint for the highest cost player, $i = 1$. Here, we have

$$\theta(z^*)(1-(1-\delta)(1+c_1(z^*)))=1.04>1$$

which is true, so all will participate. A few points are notable based on this example. First, the participation constraint means that the lower tail of the distribution of states must be "close" to the median. This means that either the variance of the distribution of state ideal points will be relatively low, or the distribution will be highly skewed to the right. Second, the optimal trade-off is made by equating weighted marginal costs with weighted marginal benefits for the median. Third, as noted earlier, if z was not public, the benefits could be greater for the states. To see this consider state 1's maximization problem. Here, it is straightforward to show that 1's optimal z would be $z_1^*=2\sqrt{10}$ which is less than z^*. Since 1 faces higher costs, it would prefer lower output for lower marginal exit costs and therefore incurs welfare losses. Fourth, it is straightforward to show that the players with lower costs get higher returns. Finally, the center gets positive utility (or rents). To see this consider the center's one-period payoff under the equilibrium. In this case, with no player shirking, it gets total income of 3 and pays out to each player $(1-\delta)(1-c_i(2))$, which summed over all of the players is 2.7, so the center gets a surplus of 0.3 in every period.

4.3 A Top-Down Model

In the previous section, the *IG*, as noted, is implicitly a model of bottom-up federalism; each state gets to have a voice in the "pre-play" determination of institutional rules. As we noted in Sections 1 and 2, however, this sequence of play does not prevail in many endogenously determined federalisms. As we will discuss below in the case of China, often times a unitary government recognizes that there are gains from *decentralization* rather than *trade*, and therefore has an incentive to *construct* sub-units to carry out these specialized tasks (Riker (1964); Qian and Weingast (1995); Alchian and Demsetz (1960); Holmstrom (1982)). In many respects, top-down and bottom-up federalisms carry with them similar design issues. In both, there is some increase in productivity through a hierarchical structure, potential moral hazard problems, and institutional design issues in order to mitigate these moral hazard problems. These similarities mean that we can modify our earlier model to examine the differences between these two types of federalism.

The critical difference between top-down federalism and bottom-up federalism is in who gets to choose the institutional rules. In terms of the *IG* and the *RG*, this translates into a change in the rule-construction stage, or the *IG*. Therefore, we introduce a slightly modified version of the *IG*, which we call the top-down institutional game (*TIG*). The *TIG* differs from the *IG* in one respect: the center chooses z^*. As before, we impose the participation constraints and allow the states to coordinate on a cut-point punishment strategy. Both of these assumptions might seem unfounded. In the case of the former, the states usually have no choice in whether they will participate or not; however, it is important to recognize that as in any agency problem, once one sets up an organization, its actors, even if created by the center, have independent interests. If the prime motivating force for compliance is force, this is consistent with our model, since the costs of exit can be interpreted as a form of such punishments. Further, if participation is mandatory, the general quality of results will be maintained; one can consider voluntary participation by the states as a base case from which we relax the assumption later. The later assumption, that the states can still coordinate on a cut-point to trigger punishments might also be unrealistic. As we argued in the previous sections, the *process of developing* a constitution, and then embodying rights and responsibilities in a transparent communication, is a critical coordination device for bottom-up federalism. In this case, again, however, we can consider the *TIG* to be a base case

from which we can relax these assumptions. Indeed, as the analysis in Section 4 indicates, the result if we remove this stage is clear: it increases the set of possible equilibria in the later stages of the game. Thus, the sequence of moves is that the states vote by majority rule on $x^{C_{TIG}}$, and the center chooses z^* subject to the participation constraints.

The *RG* is the same under both the top-down and bottom-up form. As before, the production technology means that the outputs are greater than the inputs, although the interpretation, as noted before, of what these gains are from is slightly different. Also as before, there are exit costs to potential secessionists.

PROPOSITION 4.2. *Suppose the conditions from Proposition 4.1a and 1b hold. Then under top-down federalism, weakly greater rents accrue to the center, total social welfare will be lower and z^* will be the solution to the highest cost member's participation constraint.*

Although the formal proof of this proposition is contained in the Appendix, its general outline is helpful in providing an intuition for this result. In general, the center wants to have as much institutional power z as possible. The higher the exit costs, the more the rents from a federation accrue to the center. But as exit costs get prohibitive, the states' incentives to participate decline, going to zero or even negative. When exit costs are extremely high, the center cannot commit to limited extraction, and the federalism will not congeal. So ultimately, unlike in some cases in the *IG* for the median state, the participation constraint *always binds* C's choice of z in the *TIG*. Which state is binding? Since benefits are decreasing in the functions $c(z)$, the highest cost state will be the first to drop out as C chooses a higher and higher z. Thus, C will choose the z that satisfies the participation constraint of the high cost player; that is the highest possible z it can obtain without violating the constraint. This will always be at least as high as the optimal choice of the median and sometimes higher, meaning that the center will always do at least as well as under a bottom-up system.

This result has a number of implications. First, it means that giving the center a hand in the design of institutions, not surprisingly, *shifts power to the center*. Further, this implies that a top-down system, in comparison to a bottom-up one, will always involve a weakly *greater deadweight loss to society from hierarchy*.

Second, Proposition 4.2 highlights a point made more subtly earlier. The center in a federal system has an interest in *maximizing exit costs.* By making it hard for the states to leave, it can improve its power; left with no alternative, the states' bargaining position vis-a-vis the center is drastically reduced. Two analogies from economics help to amplify this point. In the industrial organizations literature, a basic premise in imperfectly competitive markets is that incumbents try to maximize barriers to entry. The reason they do this is to limit the outside alternatives of their customers, giving them monopoly power to provide goods (e.g. Caves (197?); Tirole (199?); Porter (198?)). In the case of federalism, the center acts in a similar way. By raising the barriers to exit, it means that the center reduces potential "competition" (including state self-governance) and is able to use this power to extract "monopoly rents." A similar lesson is learned from the transaction cost literature. One of the prime concerns in this literature is with the *hold-up* problem; because of *asset specificity*—the dependence of one party's assets on another party's cooperation—the dependent party has greater rents extracted from it (Williamson (1975, 1985); Milgrom and Roberts (1991?). In the same way, if exit costs are high for sub-national units, they are more dependent on the resources of the center and therefore willing to live with a worse federal bargain.

Third, the center's creation of a federal system does not mean that it automatically can get its optimal result; it does not extract all of the gains from decentralization—some accrue to the states and some are lost. Once the center establishes a federal structure, the previously non-existent actors, the states, now become *independent players.* Further, because the design of fully-extracting institutions is hampered by moral hazard and exit problems, it means that once created under a set of rules, the states will not be any more beholden, holding those rules constant, than in a bottom-up structure.

Finally, the center's ability to extract rents increases as the distribution of state cost functions converges. The reason for this is that as noted earlier, the high-cost state's participation constraint, and thus cost function, determines the optimal choice of z by C in the *TIG.* What this means is that for a given average cost function, the center is better off the lower the highest cost player's cost function is. As the distribution of cost functions converges to a degenerate case, the greater will be z^* and thus the rents extracted by C.

5. Applying the model

The models presented above yield a number of predictions about how federalisms will be designed, how rents will be divided, and when they will be stable, sustainable institutions. In this section, we apply our approach to three cases — the development of the American Constitution from the Articles of Confederation, the Nullification Crisis a generation later, and the development of modern Chinese federalism over the last twenty years. We also use our approach to provide some comments about federalism in India, Latin America, and Russia. Although the cases do not constitute conclusive evidence, all demonstrate the plausibility of the theoretical results and point to some future extensions.

5.1 American Federalism: From Articles of Confederation to Constitution

It is possible to sketch nearly all the major American turning points from the perspective of federalism. Federalism is central to the revolutionary crisis, the debate over the Constitution, the Civil War, Reconstruction and its end, the New Deal, and the rise of regulatory state in the 1960's and 1970's.

Central to our models is the tradeoff between federal power to provide (semi-) public goods and enforcement and power to encroach on state sovereignty. If this balance is not struck properly, a federalism will stray from the course intended. If the center's power is too great, the federalism will fail because there will be over-extraction by the national government; if the center is too weak, federalism will fail because the center will under provide public goods, states will shirk, and the federalism will break down. We see both of these results in operation in the development of American federalism.

The principal criticism of the Articles of Confederation by Federalist leaders was that the national government could not supply critical public goods, primarily defense against English and European security threats, but also the maintenance of public economic structures such as a common market and common, stable currency. One of the core debates between the Federalists and Anti-Federalists concerned how to provide these goods. The Federalists believed that the national government should be granted strong taxation powers in order to have resources to achieve these ends. Some Anti-Federalists admitted a concern about the

under supply of public goods. Nonetheless, most Anti-Federalists felt that the Federalist 'solution' -- granting the national government strong taxation and monetary powers -- presented too great a risk of predation (Madison and Hamilton (178?)). In terms of our model, this debate concerned different views about how to tradeoff the center's powers to provide public goods and the risk of encroachment by the national government.

Under the Articles of Confederation, the Anti-Federalists' political power allowed them to keep the balance in their favor. Congress would pass defense bills, but they could not raise money for an army. Instead they would pass laws requiring states to provide taxes. But as the models highlight, because the center had insufficient enforcement powers, many states refused to contribute. Similarly, control of currency was also impossible. Rhode Island, for example, refused to discontinue their practice of "over supplying" and thus devaluing currency, ultimately making it difficult for the center to maintain economic property and asset values elsewhere. Further, some states hindered the development of a common market by establishing internal trade barriers, which had *not* been characteristic under British colonial rule.

In all of these cases, the central government attempted to intervene. To resolve the under supply of public goods, the Federalists consistently proposed to grant the national government taxation authority. The Anti-Federalists successfully opposed all of these initiatives, however, arguing that it would mean loss of control over the national government and hence a loss of liberty. Three times during the 1780's, the Federalists made such proposals, and all three times they were defeated, culminating with Rhode Island's veto in 1786. The veto structure of decision-making under the Articles -- in which single states could block passage of national programs -- implied that the national government had insufficient power to provide goods and enforce contributions. As our models predict, the result was classic free-riding and under provision of national goods. Our model also suggests that one of the main problems with the Articles was that they did not clearly define the limits of federal authority. When faced with granting taxation power, the Articles made the authority grant (z in the model) discontinuous: if it allowed the center to enforce taxation power, there were no limits to how far this power could be taken. The national government could either have no taxation authority or unfettered such taxation power. Fearing predation, Anti-Federalists

blocked Federalist initiatives to increase national power, the resulting in an ineffectual federalism from 1781 to 1789.

The genius of the new Constitution was in the way it resolved this dilemma through institutional rules. First, it granted the national government sufficient power to provide the critical national public goods of national defense, common markets and common currency. Second, it created limits on the national government.

Limits on the national government took several forms. First, the Constitution contained explicit limits on the national government: the national government had solely enumerated powers, with all other policy jurisdictions reserved for the states; the separation of powers system meant it would be hard for extremists to take control of the national apparatus, "ambition would check ambition;" (cite: Federalists); this system was reinforced by having an institution, the Senate, which would represent the each states directly; similarly, the Supreme Court was established with the authority to enforce these rules.[7] Second, the debates during the Revolutionary crisis and over the Constitution helped forge a consensus about how to limit the national government (Wood 1969, others). In terms of our model, the Constitution helped set limits on the national government by creating a coordination device about trigger strategies. If the national government over-stepped these limits, states would threaten to secede.[8]

For example, under the leadership of President John Adams and Secretary of the Treasury Alexander Hamilton, Federalists sought to expand the national government beyond the powers defined by the Constitution. [give example of stronger powers sought.] At the same time, the Adams administration attempted to suppress its political opposition, including the jailing of opposition newspaper editors — behavior we tend to associate more with modern Latin American states than the United States. In combination, these policies and behavior prompted a political backlash. Indeed, these attempts fostered the development of an opposition party and spring-boarded it into power in the twenty years following the election of Thomas

---

[7]As Bednar, et. al. (1995) sShow, the Court was better at policing the states than the national government.

[8]As Arthur Schlesinger, Sr., (1922) showed, every state discussed secession at one point prior to the Civil War.

Jefferson's defeat of Adams in the presidential election of 1800. The consensus lasted another generation, making the limits on government *self-enforcing*: politicians avoided violating widely-held precepts, since such violations would risk officials' political futures (Weingast (1997a)).

5.2 The Nullification Crisis (1828-1833)

Although the nullification crisis unfolded during the first Administration of President Andrew Jackson (1829-1833), it had its roots in the demise of the previous consensus established with Jefferson's election. As with the controversy over creating the American Constitution, that surrounding the nullification crisis focused on the appropriate bounds, both upper and lower, of national government power.

To understand the genesis of the nullification crisis, we begin with the crisis over the admission of Missouri in 1819-1820. The crisis demonstrated to many Southerners that their "property and institutions"— particularly through national encroachments on slavery and economic tariff policies — were not safe within the Union. They believed that, if given the power, opportunistic Northerners would attack slavery as a means of breaking apart majority coalitions and of extracting benefits from Southerners (cite).

In the period preceding the Missouri Compromise, rough parity had held between the North and the South. Many Northerners felt that the attempt to admit Missouri without a balancing free state would allow the South to gain an upper hand in control over the national government. Northerners feared Southern dominance; their weapon to protect themselves was an antislavery initiative. Northerners responded by amending the Missouri statehood bill in the House, where they held a majority, by prohibiting the import of new slaves and providing for the gradual emancipation of slaves already in residence. These amendments prompted a Senate veto and an ensuing crisis.

The Compromise of 1820 did three things to resolve the crisis: immediately, it balanced the admission of Missouri by carved off the northern counties of New England to establish the free state of Maine; for the long term, it established the 30-36 line, which divided up the remaining national territories between North and South, free and slave; finally, it

established that states would be admitted in pairs. As under the Articles of Confederation, the fundamental concern of politicians was how to design mechanisms that would allow continued operation of effective national government, but would prevent encroachment on state and local politics.

Despite the imposition of the balance rule, many radical Southerners still feared the designs of the North. During the first Jackson administration, radical Southerners in South Carolina developed a new check on national authority. Under Calhoun's leadership, they proposed the nullification a doctrine, a variant of proposals offered by Jefferson and Madison during the Adams administration, for example, in response to the 1798 Alien and Sedition Acts. Nullifiers argued that a state could interpret and defend the Constitution on its own, therefore affording it the power to "nullify" or set aside national legislation. Calhoun further claimed that the Constitution was not a "forever pact," but a compact among sovereign states which could decide to exit. The nullification doctrine meant that states could pick and choose which national legislation they would abide by, which national legislation would become law within that state. In its most clear manifestation, South Carolina responded to the dispute over tariffs during this period by nullifying the national law.

In practice, the nullification doctrine would have had two effects. First, it would have undermined the Constitution. Granting each state a veto over national policy within their borders would have crippled the national government's powers. In terms of the model, had nullification been upheld, it would have meant the dissolution of American federalism: by eliminating the ability of the central government to impose and police standards, the result would have been free-riding and breakdown. Second, nullification would have drastically lowered state exit costs: indeed, its titular purpose was to allow costless exit.

In this sense, Southern incentives reflect those studied in the repeated game of Section 4. In particular, there we make two important points. First, high exit costs shift power away from individual states and towards the national government. Second, differentials in exit costs redistribute benefits from high-cost states to low-cost states. Both of these effects appear to have motivated radical Southerners. Nullification was an explicit attempt to reduce exit costs for the South, giving them a higher degree of power against national encroachments. Further

nullification would have decreased the importance of exit costs between the states, since lower institutional power for the center automatically reduces the spread between states.[9] This too would have reduced Southern concerns about dependence and dominance by the North, since, *relatively*, their power would have been equilibrated.

The means by which Jackson and his political advisor and organizational genius, Martin Van Buren, defeated Southerners attempting to implement the nullification doctrine is also instructive. Jackson helped forge a near national consensus over a new definition of states' rights. The new definition held that the national government had virtually no role in regulating the economy, except through taxation to provide enumerated public goods and monetary policy. The advantage to Southerners was obvious: an absence of any mechanism allowing the national government to interfere with slavery. Many Northerners who also feared an overweening, remote national government, though in smaller proportion to Southerners, supported Jackson's move.

The new consensus over states' rights helped create new, self-enforcing limits on the national government. It gave the Democratic party a comparative advantage in electoral competition in the South, while allowing it to be competitive in the North. This had two related effects. First, it enabled Democrats to become the hegemonic party during the era, dominating politics from the election of Jackson to that of Lincoln. As table 5.1 reveals, Democrats held united control of the national government in 8 of the 16 Congress between the election of Jackson and Lincoln; their political opponents, the Whigs, did so in only one of 16 Congress. National policy therefore had a decided Democratic cast during this era. Second, as long as these doctrines maintained their dominant position, Democrats had no incentive to alter them. The Democrats hegemony combined with the near national consensus on states' rights to protect most Southerners and many Northerners, and conditioned the ability of national, election-seeking politicians to encroach on state sovereignty.

---

[9] To see this, consider a family of functions $c_i(z)$ such that $c_i(z) \propto c_j(z)$. Then, since the cost functions are inccreasing and convex, this implies that $|c_i(z_1) - c_j(z_1)| < |c_i(z_2) - c_j(z_2)|$ if $z_2 > z_1$. This implies that for a set of n states with such a family of cost functions, the *variance of costs is increasing in z.*

Table 5.1: Democratic Hegemony over National Elections, 1828-60.

Year	Congress	House	Senate	President
Second party system:				
1829-31	21	D	D	D
1831-33	22	D	D	D
1833-35	23	D	W	D
1835-37	24	D	D	D
1837-39	25	D	D	D
1839-41	26	D	D	D
1841-43	27	W	W	W
1843-45	28	D	W	W
1845-47	29	D	D	D
1847-49	30	W	D	D
1849-51	31	D	D	W
The 1850s:				
1851-53	32	D	D	W
1853-55	33	D	D	D
1855-57	34	W	D	D
1857-59	35	D	D	D
1859-61	36	W[1]	D	D

Source: Austin (1986), Burnham (1955), and Martis (1990)

Notes: D = Jacksonians and Democrats
W = Whigs/oppositions/Free Soilers/Republicans
* No party holds a majority, but a Republican elected speaker.

Although not all of this could have been foreseen in 1833 during the nullification crisis, Van Buren and Jackson's solution gave Southerners almost everything they wanted, except for the radical tool of nullification. In equilibrium, this tool was unnecessary. The Jackson-Van Buren approach simultaneously averted the crisis by defeating nullification, created a new, hegemonic party, provided the basis for self-enforcing limits on the national government, and thus preserved a stable federalism.

Unfortunately, this consensus was to fall apart a generation later, but we leave that tale for another time (see, however, Weingast (1997a) for a discussion of how the federalism of Jackson during the second party system broke down in the 1850s).

5.3 Modern China

Mao's death in 1976 left China in disarray. Mao's cultural revolution had been an economic and political disaster. Further, his death created a succession crisis. The latter was resolved in 1978 when Deng Xiaoping emerged as China's new leader. Deng sought to solve China's economic problems through market reform.

Potential problems of predation and opportunism were a major impediment to the central government's fostering markets. Deng addressed these problems through several strategies. First, reform was gradual, beginning with experiments that were expanded if successful and abandoned if not. Second, Deng began with agrarian reform, abandoning the disastrous collectivist system. By turning land, equipment, and other capital over to the peasants, Deng created several hundred million peasant constituents favoring reform. The result was a significant boost in peasant incomes and in total production (cites). Third, economic reform was accompanied by striking political reform. Although the Communist Party of China (CCP) retained its lock on national power, the central government devolved considerable power to lower governments. This new system of federalism granted considerable autonomy and power to the provinces and lower governments (Montinola, Qian and Weingast (1995)).

Agrarian reform contributed to the central government's commitment to economic reform in three ways. First, it created a huge, pro-reform constituency. Second, this could be undone only at the price of massive violence against the peasantry. Third, it demonstrated to others that the central government's new initiative were not tentative.

By the mid-1980s, China sought to extend its reform to industry and commerce. Here too, the problem of central government predation and opportunism loomed as a large impediment, since fears of such encroachment would vastly increase the uncertainty related to capitalist investments. The central government sought to limit the possibilities of predation and

opportunism in several ways. First, it devolved considerable power to the lower governments. Central to this devolution were new fiscal powers. Local governments, not the national government, collected taxes, forwarding the national government an agreed amount and keeping the residual. Local governments were also granted regulatory authority over the economies. These governments, not the national government, became the locus of decision-making over rules governing production and exchange. Finally, the national government slowly dismantled its planning and spy apparatus.

These institutional changes had several effects. The new fiscal powers allowed lower governments to act as residual claimants to their economies. Because they could keep tax revenue beyond a certain amount, they had strong incentives to foster local economic prosperity. Economic growth would benefit local citizens and local governments, not the national government. Although not all local governments initially followed this path, several on the south coast did so aggressively, particularly Guangdong province. As Guangdong's impressive success became apparent, other provinces and localities began to imitate it.

At the same time, fiscal reform also limited the national government's resources in unforeseen ways, making it the poor relative to its political obligations (e.g., its welfare obligations associated with the SOEs). Importantly, fiscal stress further limited the central government's ability to encroach on the provinces.

The dismantling of the planning and spy apparatus also reduced the threat of encroachment. As economists emphasize with respect to the socialist planning system, the central government's information enhanced its ability to encroach and implied an inability to commit to non-interference (Milgrom and Roberts 1990) — e.g., not to raise quotas (Laffont and Tirole); not to subsidize, creating the so-called soft budget constraint (Dewatrapont and Maskin; generally, see Riordan and Aighon and Tirole). Dismantling the central government's information systems reduced its ability to extract from lower governments and firms. Indeed, the Chinese have a phrase reflecting this, "[get]", meaning "storing wealth in enterprises".

We interpret China's policies for creating economic reform as including political reform that created a new system of top-down federalism. By granting the provinces and lower

governments new powers, China created a set of political actors with incentives to resist national encroachments on lower government power. All governments had incentives to resist.

Events after the bloody suppression of the Tiananmen Square demonstrations illustrate this point. This period witnessed the anti-reformists' strongest moment of power within the central government during the entire reform period (1978-present). At this time, China's Premier, Li Peng, sought to undo the fiscal system and provincial autonomy. A similar move had occurred on two previous occasions under Mao; both were successful. But in 1989, at a meeting of governors of the provinces, the governor of Guangdong province said no (Shirk 1993). Because so many provincial governors sided with Guangdong's governor, Li Peng was forced to back down. China's new system of federalism survived its biggest challenge. As our model suggests, the trigger-strategy threat of non-cooperation by the provinces proved central to policing the center's willingness to adhere to federalism's rules.

Present political and economic problems facing China

Our model illuminates a range of China's present political and economic dilemmas. First, the national government has huge fiscal liabilities relative to its fiscal resources, hindering its ability to provide public goods. A major set of problems concerns the old state owned enterprises (SOE). Because they employ so many workers, the national government cannot just abandon these enterprises — that would cause far too much civil unrest. Second, there are growing disparities among regions. Preservation of public peace and internal security may well require some form of redistribution. Third, China lacks critical infrastructure public goods, such as electric power and transportation.

The optimal economic solution to these three problems is a more powerful center, one with sufficient resources to provide these public goods. Several public finance economists have proposed doing just this (see, e.g., Wong and ***). A major problem exists with this proposal, however, one ignored by the traditional neoclassical public finance approach, namely predation and opportunism. If the center is sufficiently endowed, it could also undo everything about the market. And the successful reform provinces are scared to death of this.

The Chinese have not ignored their public goods problems, however. Indeed, our model can interpret China's reaction as a second best solution to the problems of under provision of national public goods in the face of potential government opportunism. This solution reflects a "pragmatic" set of arrangements.

First, for transportation, many initiatives have been financed by wealthier provinces with specific commercial and other links to poorer ones; and with each other. This is unlikely to result in the optimal set of transportation facilities, due to free riding and the lack of an institutionalized forum within which to negotiate these bargains among the provinces. Nonetheless, they are creating a better transportation system than would result if every province shirked.

Second, instead of the richer provinces agreeing to provide the national government with greater fiscal resources to relieve its fiscal stress, they have agreed to take over some of the national government's major fiscal liabilities. For example, there has been a large transfer of responsibility for SOEs, including their welfare responsibilities, from the national to provincial governments.

Third, with respect to redistribution, a series of ad hoc relationships among provinces have emerged. For example, because many coastal provinces need large supplies of cheap labor, they have institutionalized arrangements with interior ones to regularize the flow of labor. In return, poorer provinces receive two benefits. First, laborers on the coast, often earning 10 times what they could earn back home (Solinger 19**), send back considerable remittances. Second, many laborers return after several years as individuals rich in human and liquid capital. This new capital is then used for new economic enterprises back home.

The biggest problem facing China concerns the fundamental tradeoff of federalism identified above. Because the central government lacks virtually all the traditional institutional means of commitment (for example, elections, a separation of powers system), the provinces are deeply worried about opportunism. Paralleling the United States under the Articles of Confederation, China's system of federalism remains biased against national power. The under supply of public goods is therefore a major economic and political problem.

In principle, China might mitigate some of these problems through creating new national institutions that place limits on the central government's behavior. For example, China could create something paralleling the original United States Senate, where the provinces are directly represented before the national government. This would provide the provinces with an institutionalized forum for negotiating with the central government and vetoing its decisions. In terms of the model, this institution would allow the provinces to collude against the center; transfer rents from the center to the provinces; and finance the national government's provision of public goods while policing the national government's ability to act opportunistically. This would also limit the center's ability to act opportunistically. Our model also suggests why China has not created such an institution. Despite the greater gains from cooperation, creating an institution to represent the provinces would lower the national government's share of rents and place significant constraints on the CCP, something Chinese leaders have not yet been willing to do.

In sum, China remains in a struggle along the fundamental tradeoff of federalism.

5.4 The Perils of Too Much Power: Overarching Centers and Failed Federalisms

Although India and many of the large Latin American states (Argentina, Brazil, Mexico) are nominally federal systems, they are degenerate federalisms. All exhibit the problems of too strong a center. In Latin American federalisms, for example, the central government provides states with 80-90 percent (or more) of their revenue. Along with the revenue comes the national government's rules and restrictions. Further, in all these states, the national government, not the local government, remains the locus of regulatory control over the economy. Any attempt to ignore the rules risks the withdrawal of all funds. For example, consider the rise of political competition to the PRI, the political party that has dominated Mexico since their Revolution. The national government frequently removes all financing from local governments captured by the political competition. This sets a very high price on voting for the opposition for local citizens.

In combination, these features of federalism imply that states are not autonomous, sovereign entities, as federalism requires. Instead, they remain administrative agents of the

national government. Paralleling the obvious implication of these rules, our model suggests that the national government captures the lion's share of the rents in these systems.

Russia reflects a different variant on the same problem. In the former Soviet Union, the central government dominated all political decision-making. Local governments at all levels were administrative agents of the central government. Moreover, the parallel party and police state apparatus implied significant punishments for individual politicians who might steer an independent course.

This legacy sets the stage for the political conflict in modern Russia. Having succeeded the Soviet Union for sovereignty over its territory, Russia has the problem of too strong a center. The Russian government today has only a limited ability to commit to limits on its behavior. In particular, states lack a consensus about the appropriate limits on the national government.

Further, the central government is financially weak. This has allowed many local and regional governments to grab considerable *de facto* independence. Short of announcing sovereignty, as in Chechnya, the central government has acquiesced to most of these assertions of regional government power.

Our model provides some important insights into Russia's current problems. All regions and the national government would be better off under a clearer definition of federalism, but the states and the central government differ about the distribution of the rents. In principle, the center could foster the state's cooperation, but that would transfer most rents to the states. In practice, the central government has attempted to deal with the states one at a time, isolating them and preventing their collusion. Solnick (1996) interprets this behavior by the center using the chain-store model: bargaining one-on-one with the states instead of all of them at once grants the national government greater bargaining leverage, since the national government can potentially overwhelm any one state. States, on the other hand, resist solutions to the federal problem that grant most of the rents to the center. Because neither the center nor the states are sufficiently powerful to win this contest, there remains a standoff.

Notice also that Russia's assault on Chechnya reflects the issue of exit costs. As our model suggests, the rents accruing to the center are a positive function of exit costs. Allowing

Chechnya to leave without a fight would have signaled a lower exit cost to other regions. Given the uncertainty over the future of Russia's structure, the likely result would have been many more states resisting the center's policies; perhaps many more would have sought to leave; all suggesting a much worse bargain for the center. As with the cases in India and Latin America, the Russian case demonstrates that when the balance between central and state authority is struck inappropriately, federalisms will become degenerate and ineffectual.

6. Conclusions

We began our study with twin dilemmas of federalism: too strong a center risks overwhelming federalism by acting opportunistically and extracting too many rents; too weak a center risks federalism's collapse due to free-riding. The twin dilemmas make stable federalism problematic, in part because they imply a tradeoff in the structure of federalism. Institutions designed to address one of the dilemmas risks exacerbating the other. To be stable, therefore, federalism requires a delicate balance of central government powers combined with mechanisms for limiting the center's opportunism.

Our models suggest a series of results. First, for federalism to overcome the shirking problem, states must be sufficiently fearful of the center. This means that the center must be imbued with sufficient monitoring resources and penalizing capacity. Second, the benefits from federalism must be sufficiently large so that both the center will not "take the money and run", expropriating all contributions, and the states will be better off. Third, when states can *coordinate* on punishment strategies, this will lead to greater provision of public goods, and higher public welfare. Fourth, the introduction of *exit costs* shifts rents back to the center. The logic of this point is that as the states' costs of exiting increase, their threat to exit becomes less credible. This increases the bargaining power of thee center against the states, and shifts some of the rents to the center. Fifth, heterogeneity among the states means that some end up subsidizing the center more than others. In particular, those states who face very high costs of exit receive fewer benefits to federalism. Since adequate incentives have to be provided for the center to continue to provide public goods, these incentives will be disproportionately provided by the relatively weaker states. Finally, in choosing the optimal amount of institutional power granted to the center, designers can effectively resolve the two dilemmas. This resolution leads to a level

of public goods provision that is less than would be socially desirable. Further, if the inappropriate level of institutional power is given the center, it will be destabilizing.

An important feature of our approach is that states' ability to coordinate is critical to resolving the dilemma of central government encroachment and opportunism. The creation of a constitution, for example, serves to construct a focal point coordinating state reactions against a central government that seeks to violate the rules. Thus, as many observers of federalism suggest, there might appear to be a "culture of federalism" helping sustain successful federalisms (Elazar 19**). We differ with these scholars over one critical point. They typically see culture as exogenous: only those federal states with such a culture survive. Our approach instead suggests that this culture is endogenous, a product of the design stage. The two episodes described in the United States' history — the creation of the Constitution and the redefinition of states' rights under Andrew Jackson during the nullification controversy — both exhibit the construction of a set of consensus agreements about the limits on the national government and on state shirking. In this view, the construction of a coordination device helps create a "federal culture" and sustain federalism.

Our approach also suggests an important difference between top-down and bottom-up federalisms. Federalism designed by the center is likely to leave the center with a greater share of the rents, while bottom up federalism is likely to leave the center with the minimum level of rents necessary to perform its functions. Further, this means that *ceteris paribus*, that top-down federalisms will be *weakly more socially inefficient* than bottom-up federalisms. Since greater rents to the center mean lower public goods provision, the additional power granted to the center in bottom up federalism will lead to dead-weight losses.

In Section 5, we demonstrated both the strengths and weaknesses of our theoretical results through the examination of a number of cases. In all of the cases, the potential to gain the benefits from cooperation and public goods provision was traded against the difficulties of shirking and encroachment. In each case, the result was clear delimitation of central power. But in delimiting such power, these institutions also provided the central government with authority *to that point.*

Although incomplete, the theoretical developments in this paper point to important directions for future research. Returning to the stylized issues introduced in Section 1, we can

reexamine our theory in light of the cases and these facts. The cases suggest an important extension to the theoretical model. We model the fundamental trade-off as one in which the ability to provide public goods (in the form of θ) is traded against potential encroachment (which increases in the c's). Another, perhaps more empirically general, way to model the trade-off is to endogenize the center's enforcement powers, as represented in f. This would mean that increases in central power would come with higher incentives for the center to employ "unfair" enforcement strategies, fining those who had not shirked. The states would then face a choice of having a powerful, encroaching center or losing some of the benefits to shirkers.

Although this paper does not resolve all of the issues, it demonstrates the power of what Gibbons and Rutten have referred to as the "equilibrium institutions" perspective. For students of federalism, our approach demonstrates the power of such a perspective. Using the formal tools of rational choice and institutional analysis, we have attempted to focus attention on the specific trade-offs and requirements of stable federal institutional arrangements. More generally, for students of constitutions and democratic institutions, we use the *case* of federalism to demonstrate how one might think about constitutions themselves. In the vast majority of the literature examining institutions, these rules are taken as exogenous: we among others have argued that such rules help to shape behavior in the political arena. In this paper we move closer to Gibbons and Rutten's ideal (Gibbons and Rutten (1996); Diermeier (1995)). By taking the approach that constitutions should be studied as self-enforcing equilibria, in other words as *endogenous*, we have demonstrated not only the force of such documents but also their rationales.

APPENDIX: PROOFS OF PROPOSITIONS STATED IN TEXT

Proof of Proposition 2.1. Since the payoffs are bounded and there is discounting, we can use the Optimality Principle of dynamic programming to check only single period deviations. In this case, if a player deviates from the cooperative path, he gets the Nash result in every period after, so he gets $\frac{\theta}{n}(n-1)$. If he cooperates he gets $\theta-1$ in every period, which is $\frac{\theta-1}{1-\delta}$ in the repeated game. Thus, player *I* will cooperate if the former quantity is less than the latter one. The result follows by rearranging this inequality.

Proof of Proposition 2.2. For δ, this follows directly from (2.1). For the rest of the proposition, it is sufficient to show that δ^* in (2.1) is increasing in n and decreasing in θ. (I) n: $\frac{\partial\delta^*}{\partial n}=\frac{\theta(\theta-1)}{(\theta n-\theta)^2}$ which is positive since by assumption, $\theta>1$. (ii) θ: $\frac{\partial\delta^*}{\partial\theta}=\frac{-n(n-1)}{(\theta n-\theta)^2}$ which is clearly negative.

Proof of Proposition 3.1a. Let x be the *C's* offer. Fix z, δ. Consider first the states' strategies. In every period they cooperate, a typical state will get $u_i=\theta x-1$. In periods in which they shirk, a state's expected payoff is $u_i=\theta x-fq$. Thus, a state will only shirk if $fq < 1$, which is ruled out by assumption. Now consider what happens if a state plays *E*, when the center is playing its equilibrium strategy. In this case, the state will exit iff $\sum_t \delta^t(\theta x-1)<0$ which will only hold if $x<\frac{1}{\theta}$ which is also ruled out by assumption. So any equilibrium offer greater than this quantity can be supported as an equilibrium with trigger strategy x. Now consider *C*'s strategy. Suppose the states' are playing a cut point trigger strategy of x. It will be sufficient to show the values of x for which *C* will not deviate. Here, we have the center gets $\sum_t \delta^t(1-x)$ for cooperating. If it deviates, the center gets n. Here it is sufficient to consider a single state's payment. So the *C* will deviate iff $x>\delta$.

Proof of Proposition 3.2. Here, the states can coordinate on any equilibrium punishment strategy. They will choose the one which generates the most welfare for each individual state. As shown above, the center will not deviate if $x<\delta$, and the states will cooperate for any $x>\frac{1}{\theta}$. It is clear then, that the states will choose the upper bound and can coordinate to punish

any payments less than $x^C=\delta$. Finally, since this is a dominant strategy for all players, then all will choose to vote for it over any other proposal and therefore x^{C*} is an element of the core.

Proof of Proposition 4.1a. Let x be the *C's* offer. Fix z, δ. Consider first the states' strategies. In every period they cooperate, a typical state will get $u_i=\theta x-1$. In periods in which they shirk, a state's expected payoff is $u_i=\theta x-fq$. Thus, a state will only shirk if $fq < 1$, which is ruled out by assumption. Now consider what happens if a state plays E, when the center is playing its equilibrium strategy. In this case, the state will exit iff $\sum_t \delta^t(\theta x-1)<-c_i$ which will only hold if $x_i<\frac{1-c_i(z)(1-\delta)}{\theta}$ which is also ruled out by assumption. So any equilibrium offer greater than this quantity can be supported as an equilibrium with trigger strategy x. Now consider C's strategy. Suppose the states' are playing a cut point trigger strategy of x. It will be sufficient to show the values of x for which C will not deviate. Here, we have the center gets $\sum_t \delta^t(1-x_i)$ for cooperating. If it deviates, the center gets $\sum_t 1+\delta c_i$. Here it is sufficient to consider a single state's payment. So the C will deviate iff $x_i>1-(1+\delta c_i)(1-\delta)$.

Proof of Proposition 4.1b. *(i)* Here, the states can coordinate on any equilibrium punishment strategy. They will choose the one which generates the most welfare for each individual state. As shown above, the center will not deviate if $x_i<1-(1+\delta c_i)(1-\delta)$, and the states will cooperate for any $x_i>\frac{1-c_i(z)(1-\delta)}{\theta}$. It is clear then, that the states will choose the upper bound and can coordinate on any payments less than $x_i=1-(1+\delta c_i(z))(1-\delta)$. Finally, since this is a dominant strategy for all players, then all will choose to vote for it over any other proposal and therefore x_i^{C*} is an element of the core. *(ii)* The states will maximize their ongoing payoffs which is equivalent in this case, to maximizing the payoffs in a single stage. Given that a state will get $\theta(z)(1-(1-\delta)(1+\delta c_i(z)))-1$ in every period on the equilibrium path, then the solution to each *individual state's* problem is the solution to

$$\max_z \quad \theta(z)(1-(1-\delta)(1+\delta c_i(z)))-1$$

Solving the first order conditions, we obtain an implicit *ideal point* z_i^*, which is the solution z to condition (*). That these are maxima, we check the second order conditions, and obtain

$$\frac{\partial^2 \varphi}{\partial z^2} = \theta''(1-c) - c'\theta' - c''\theta - c'\theta'$$

where we have suppressed the arguments z and the subscripts i since they are clear. Analyzing the second order condition, we have since $\theta'' < 0$ and $1\text{-}c > 0$, the first term is negative; since c', $\theta' > 0$, the second term is negative; since c'', $\theta'' > 0$, the third term is negative; and c', $\theta' > 0$, the last term is negative; which implies that the objective function is concave and the condition (*) defines a maximum. Now, we must check which proposals are in the core; in other words which proposals beat all other proposals. As mentioned, the solutions to (*) implicitly define a set o f ideal points in a unidimensional z space. Thus, the median voter theorem of Black holds in this case. However, we must also show that the participation constraint is satisfied. Order the z_i's such that $z_1 \le z_2 \le \ldots \le z_N$. Then the participation constraint will be satisfied for $\forall i$ iff $\theta(z_M)(1-\delta)(1-c_1(z_M)) - 1 \ge 0$, which is true by assumption.

Proof of Proposition 4.2. We can use the solutions to propositions 4.1a and 4.1b in every respect except one. Here we must solve *C's* problem which is

$$\max_z \quad (1-\delta)\sum_i (1+\delta c_i(z)) \qquad \textit{s.t. } N \textit{ participation constraints}$$

It is clear from this maximization problem that, for the unconstrained problem, there is no well-defined solution: *C* will choose $z \to \infty$. The constraint puts a bound on how high z can be. From the proof of proposition 4.1b, we know that the binding member is the state with the highest cost function, since this member wants the lowest z. Thus, the z which solves *C*'s constrained problem is that which solves the equality condition in the high-cost member's participation constraint. The fact that the rents to the center are weakly greater follows from the fact that if the participation constraint in the *IG* is not binding, then z will be lower than

when the constraint binds. In this case, the constraint always binds, meaning $z^*_{TIG} \geq z^*_{IG}$. The fact that this means higher rents follows by noting that C's utility is increasing in z.

References (incomplete)

Axelrod, Robert. 1984. *The Evolution of Cooperation.*
Bednar, Jennifer L. 1994. "The Federal Problem, MS, Stanford University.
Bednar, Jennifer L. 1996. "Federalisms; Unstable by Design," MS, Stanford University.
Bednar 1997
Bednar, Eskridge, and Ferejohn. 1996. "A Political Theory of Federalism."
Bednar, Jennifer L., John A. Ferejohn and Geoff Garrett. 1996. "The Politics of European Federalism," *International Review of Law and Economics*, pp.279-294.
Bendor, Jonathan, and Dilip Mookherjee. 1987. "Institutional Structure and the Logic of Ongoing Collective Action," *American Political Science Review,* pp. 131-154.
Calvert Randall L. 1992. "Rational Actors, Equilibrium, and Social Institutions," MS.
Caves, Richard E. 1972. *American Industry: Structure, Conduct, Performance.*
Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory.*
Gibbons, Robert, and Andrew Rutten. 1996. "Hierarchical Dilemmas: Social Contracts with Self-Interested Rulers," MS.
Green, Edward, and Robert Porter. 1984. "Noncooperative Collusion Under Imperfect Price Information." *Econometrica*, pp. 87-100.
Greif, Avner, Paul Milgrom, and Barry R. Weingast. 1994. "The Merchant Guild," *JPE*
Kreps, David. 1990. *A Course in Microseconomic Theory.*
Kreps, David M., and Robert B. Wilson. 1990. "Reputation and Imperfect Information," in Oliver Williamson, ed., *Industrial Organization.*
Milgrom, Paul R., Douglass North, and Barry R. Weingast. 1990. "The Law Merchant," *Economics and Politics*
Milgrom, Paul R., and John Roberts. 1990. "Bargaining and Influence Costs" in James Alt and Kenneth A Shepsle, eds., *Perspectives on Positive Political Economy.*
Montinola, Gabriella, Yingyi Qian, and Barry R. Weingast. 1995. "Federalism, Chinese Style" *World Politics*
Oates, Wallace. 1992. *Fiscal Federalism.*
Ordeshook, Peter. 1996.
Osborne. Martin J., and Ariel Rubenstien. 1994. *A Course in Game Theory.*
Persson, Torsten, and Guido Tabellini. 1996. "Federal Fiscal Constitutions: Risk Sharing and Moral Hazard," *Econometrica*, pp. 623-646.
Porter, Michael E. 1980. *Competitive Strategy : Techniques for Analyzing Industries and Competitors.*
Riker, William H. 1964. *Federalism.*
Rubinfeld, Daniel. 1987. "Economics of the Local Public Sector," in A. J. Auerbach and M. Feldstein, eds., *Handbook of Public Economics* vol. II.U
Shirk, Susan. 1993. *The Political Logic of Economic Reform in China.*
Solinger, Dorothy. "The Floating Population as a Form of Civil Society," Paper for 43d Annual Meeting of the Association for Asian Studies, New Orleans, April 11-14, 1991.
Tirole, Jean. 1988. *The Theory of Industrial Organization.*

Treisman, Daniel. 1996. "Crises and Stability in Federal States: A Game-Theoretic Analysis," MS, Russian Research Center, Harvard University.

Weingast, Barry R. 1995. "The Economic Role of Political Institutions: Market-preserving Federalism and Economic Development," *Journal of Law, Economics, and Organization.*

Weingast, Barry R. 1997a. Civil War MS.

Weingast, Barry R. 1997b. "The Political Foundations of Democracy and the Rule of Law," *American Political Science Review.*