Self-Enforcing Federalism

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

First Version: May 1997
This Version: March 2002

ABSTRACT

How are constitutional rules sustained? We investigate this problem in the context of how the institutions of federalism are sustained. As Riker (1964) emphasizes, a central design problem of federalism is how to create institutions that at once grant the central government enough authority to provide central goods and police the sub-units, but not so much that it usurps all public authority. Using a game theoretic model of institutional choice, we argue that, to survive, federal structures must be *self-enforcing*: the center and the states must have incentives to fulfill their obligations within the limits of federal bargains. Our model investigates the tradeoffs among the benefits from central goods provision, the ability of the center to impose penalties for non-compliance, and the costs of states to exit. We also show that federal constitutions can act as coordinating devices or focal solutions that allow the units to coordinate on trigger strategies in order to police the center. Finally, the model generates a number of comparative statics concerning the degree of central power, the division of rents between the states and the center, and the degree of "central goods" provided, as a function of the characteristics of the constituent units.

1. Introduction

How are constitutional rules sustained? Although a long normative tradition exists about various aspects of constitutionalism, a positive literature on this topic is only just emerging.[1] The general problem concerns how to structure the political game so that all the players – elected officials, the military, economic actors, and citizens – have incentives to respect the rules.

---

* Haas School of Business and Department of Political Science, University of California at Berkeley; Hoover Institution and Department of Political Science, Stanford University. The authors gratefully acknowledge the helpful comments of Jenna Bednar, Jonathan Bendor, Mel Bernstein, Scott Gehlbach, Natalia Ferretti, Ed Green, Douglas Grob, Dan Kelemen, Barak Richman, Pablo Spiller, Ken Shepsle, Oliver Williamson and seminar participants at the the Harvard University, Political Economy of European Integration working group, Princeton University, Stanford University, University of California at Berkeley, University of California at Los Angeles, University of California at San Diego, San Diego University, and Yale University.

[1] See, for example, Fearon (2000), Calvert (1995), Greif (1997,2001), Hardin (1989), Milgrom, North and Weingast (1990), Ordeshook (1992), Przeworski (1991, ch 2; 2000), and Weingast (1997).

In this paper, we investigate this problem in the context of how the institutions of federalism are sustained. We follow Riker (1964,11) and define a government as federal if it has a hierarchical governmental structure in which each level of government has some autonomy.

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

> **Dilemma 1:** What prevents the national government from destroying federalism by over-awing 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. Further, since constitutions are not externally enforced, such a resolution requires that the rules defining a federation be self-enforcing for political officials at all levels of government. Our work contributes to a new and growing literature which Gibbons and Rutten (1996) call the new "equilibrium institutionalists."[2] Scholars in this tradition observe that, for constitutional features to endure, political officials must have an incentive to abide by them.

Resolving the two dilemmas is problematic because they imply a *fundamental tradeoff*: mechanisms to mitigate one dilemma typically exacerbate the other. Too weak a national government will exhibit free-riding and insulated, "dukedom" economies. Or worse, it will disintegrate. With a national government too strong, a federation typically fails because the national government compromises state independence. Reflecting this tradeoff, several theorists emphasize federalism's instability (Riker 1964, Bednar 1996, Bednar, Eskridge, and Ferejohn 2001, and Ordeshook and Shvetsova 1997).

Three rich streams of the literature relate to the two fundamental dilemmas of federalism.[3] The first and largest stream studies the problem of state shirking and common pool problems from sub-national governments. The settings vary dramatically, including demand for federal spending;

---

[2] See, e.g., Bednar (1996, 1998a), Calvert (1996), Gibbons and Rutten (1996), Greif (1997, 2000), Greif, Milgrom, and Weingast (1994), Milgrom, North, and Weingast (1989), and Weingast (1997).

[3] These three literatures focus on aspects of endogenous federalism. In addition, there is a much larger literature on the effects of federalism, dominated by the economists (such as Oates 1972, Rubinfeld 1987, Tiebout 1956). There is also a political science literature on the effects of federalism on various problems, such as ethnic conflict (see Lijphart 1984,ch10), budget deficits (Rodden 1999, 2000), Poterba and von Hagen 2000), and corruption (Treisman 1999).

budgets, state borrowing, soft budget constraints, and deficits; and voting.[4] These scholars show that, without a strong center, common pool problems produce third-best or even worse outcomes. The focus on the common pool problem emphasizes the second dilemma of federalism, the failure of "too much" decentralization.

The second stream of literature examines the first fundamental dilemma, the problem of national government aggrandizement. Bednar (1998a, 1998b) and Riker (1964), for example, examine how central governments tend to expand their powers over time. Ordeshook and Chen (1994) study the problem of how a central government can be prevented from usurping all public authority. Weingast (1997) examines how a central authority can use a "divide and conquer" strategy to transgress its authority without reprisal (see also Treisman 2000).

Finally, a third literature examines the joint problem, albeit in very specific contexts. Riker (1964), Garman, Haggard, and Willis (1999), and Ordehsook and Shvetsova (1997) emphasize the role of parties for federal stability. They argue that the need to cooperate to win elections drives national and subnational officials to respect one another's interests. Bednar, Eskridge and Ferejohn (2001) conclude that although judicial institutions have an asymmetric effect: they tend to police the subnational governments, but are less effective in policing national government aggrandizement.

In this paper, we synthesize aspects of these literatures. The first two literatures each emphasize one of the two federal dilemmas and thus study half of the problem. Our approach studies the two problems simultaneously. Similarly, although the papers in the third literature recognize the problem we discuss here, we complement them by generalizing their examination of *specific* institutions in developing a generic model.

To understand how successful federal systems simultaneously resolve the two dilemmas, and thus provide for their stability, we begin with the rationales for constructing federal systems. Two conditions must exist for a federal system to emerge: there must exist some gains from cooperation between sub-national units, and those gains must not be available in other institutional forms.

An additional question about federalism concerns why these systems need a central structure at all. As the first stream of literature emphasizes, the answer is that participating states want central

---

[4]See, e.g., Bednar (1998a, 1998b), Bednar, Eskirdge, and Ferejohn (2001), Blanchard and Shleifer (2000), Cremer and Palfrey (1999), Inman and Fitts (1990), Inman and Rubinfeld (1997), Jones, Sanguinetti and Tomassi (1999), McKinnon (1997), Persson and Tabelinni (1996, 1996), Poterba and von Hagen (2000), Rodden (1999, 2000), and Sanguinetti (1995).

goods, yet each has an incentive to *shirk* or "free-ride." Moreover, imperfect information about shirking 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). If the moral hazard problem by the sub-units is too severe, the states will be unable to capture the gains from cooperation in a decentralized manner. A primary solution provides the center with policing authority so it can act as a central monitor in the hierarchical structure.

National governments have their own interests, however. Granting resources and powers to the central government enables it to usurp state authority and extract resources — in Riker's (1964) term, to overawe the states. 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. The example of defense makes clear the tradeoff: giving the national government greater resources allows appropriate defense against external threats; but increasing central resources also makes it harder for the states to resist encroachments by the center.

We incorporate the two dilemmas into a repeated game that captures the nature of federal arrangements. By endogenizing federal authority, state participation and shirking, and limits on the federal government, we provide insight into two aspects of the problem. First, we derive a set of sufficient conditions for a self-enforcing federal system. Second, and perhaps more importantly, we develop a number of comparative statics implications for the institutional design of federations. In particular, we propose hypotheses about when self-enforcing federations can exist, the degree of central power, the division of rents between the states and the center, and the degree of "central goods" provided, as a function of the exogenous characteristics of the constituent units.

In the next section, we describe a two-stage model of a set of states endeavoring to capture some gains from cooperation. In the first stage the states must collectively choose a set of arrangements to define how the federation will operate. In the second stage, the states and the center interact on an on-going basis within the framework they have erected. The players in the model are a set of n states and the center. We study the problem with aggregate, unitary actors for two reasons. The first is tractability. The second is that we seek to model a problem common to all federal systems, not just about ones with political institutions like the United States, where the subnational units have identifiably different interests. Thus, we wish to model the problem for United States, Mexico, Russia, China, and the European Union.

We then proceed to analyze the model in three parts. We start by assuming that both the participants and the institutions in the federation are fixed; we then consider what happens when the institutions (but not the participants) are endogenous; and then finally we turn to the question of endogenous institutions and participants. The analysis shows that if both the penalties imposed for shirking and the probability of being detected are jointly high enough, then shirking can be prevented and the gains from cooperation potentially realized. However, unlike in the previous literature on the common pool problem, the model also illustrates that once created, the central government is not a faithful, welfare maximizing agent of the states. It has incentives to capture rents.

In Section 4, we take up the question of the institutional design of federalism, including the question of how subnational units limit the center's aggrandizement. Institutional design incorporates both grants of central authority and the choice of the "trigger" strategies to be played once the federation becomes an ongoing concern. Choosing the appropriate trigger strategies allows states to *coordinate* on a punishment regime to police the center, thus ensuring maximal benefits returned to the states. This framework generates several interesting results. We show how coordinating devices, such as constitutions, can serve to minimize efficiency losses, police the center, and maximize the return of rents to the states. Finally, we show how the equilibrium strength of the center varies with the exogenous characteristics: a weaker the set of states, a more productive center, and larger federations, *ceteris paribus*, all lead to weaker central authority.

In Section 5, we extend the model to the important question of the optimal size of federations. Here we analyze the question of which states will be included in an "equilibrium federation." As in the existing literature (see, e.g., Alesina and Spolaore 1997; Alesina, Spolaore, and Wacziarg 2000; Bolton and Roland 1997), our model shows that there is a tradeoff between heterogeneity and scale. The difference, however, is that we show that this tradeoff is not only in terms of policy efficiency, but also in terms of institutional choice. Including a state in a federation that is weaker than the existing states requires diluting central power to prevent *ex post* opportunism against the weaker state by the center. The other states in the federation will therefore choose to include a marginal state only if the scale benefits from its inclusion more than offset the costs associated with dilution of central authority.

In Section 6, we describe three cases which illustrate one of the important predictions of the model. In each, the stability of a federation was threatened by an inability to resolve the two dilemmas simultaneously. And in each case, the *relaxation* of the trade-off led to federal stability.

The first two, both American examples, concern the initial design of the United States federal system in the post-Revolutionary era; and the first major challenge to that system, during the Jacksonian period, when the Nullification Crisis occurred. In the third example, we turn to modern Chinese decentralization and the institutional issues faced there. All these cases exhibit the model's central features: insufficient power at the center meant common pool problems and shirking occurred at the sub-national level, but the reason, initially, that sufficient power was not ceded to the center was precisely because of fears of central aggrandizement. In other words, the failure to define *limits* on the center led to progressive degeneration of the federal bargain, and only when such limits were defined, could sufficient authority be granted to solve the problem of cooperation and coordination through the exercise of central authority.

Our conclusions follow.

2. A Model of Bottom-up Federalism

In this section we propose a model of federalism and institutional choice. We incorporate the following features into our model: an ongoing, stable federation must be one which *repeatedly* solves the two fundamental tradeoffs; there are benefits to *scale* in a federation; there is *heterogeneity* among the subunits; there can exist *costs for exiting* from the federation; that states have a *collective* incentive for participation, but an *individual* incentive to shirk; that all players want to maximize their *lifetime rents*; and that *monitoring is imperfect*.

To model these characteristics, we posit two stages to the complete game. The first stage is called the *institutional game (IG)* in which the institutions of the federation are determined. The second stage is the *repeated game (RG)* in which the players interact repeatedly given the institutions determined in the *IG*. Our strategy, as shown in Figure 2.1, is first to solve the characteristics of the federal equilibrium *given* the institutions of the federation, and then to understand what types of institutions will be adopted given a set of states that aim to establish a central government.

Once a federation has been established, the players interact repeatedly, making decisions about the degree to which they will comply with the requirements of the original understanding. To formalize this aspect of a federation, the *RG* is the infinite repetition of the following stage game. The *RG* has $N + 1$ players, n states indexed by $i = 1,...,N$, and a *central government* called *C*. The

sequence of moves is shown in Figure 2.2. First, the states choose one of three actions $A = \{C, S, E\}$ for contribute, shirk, and exit. If a state chooses *C,* it *contributes* one unit to the center. The indicator variable k_i=1 if a state contributes and 0 if it does not. If a state chooses *S*, it *shirks* and contributes zero. The contribution by the states represents any costly actions to a state which aids the capture of public goods, such as tax payments to the center for national defense, enforcement of regulations, or enforcement of cross-border policing agreements.

States can always choose to leave the federation. Thus, if a state chooses *E*, it contributes nothing and permanently *exits* or secedes from the federal system. We designate a state's exit choice by the indicator variable s_i, which equals 1 if the state chooses to exit and zero otherwise. If a state exits, it no longer participates in the game, incurring no costs or benefits in later stages. In addition, secession can be very costly for a state that secedes. To capture this fact, when a state chooses to exit, it incurs a cost $c_i(z)$, where we use *z* to denote the institutional authority of the central government. This formulation assumes that the greater the central government's authority and resources, the greater the costs incurred by secession. Secession is costly in part because of the need to disentagle from the federation and to establish itself as an independent state. Further, the more powerful the center, the greater the costs it can impose on potential secessionists. Therefore, we assume that this cost function is an increasing, convex function of *z*, so that $c_i(z) \geq 0, c_i'(z) > 0, c_i''(z) \leq 0$. Further, these costs differ across states: greater economic power or being on the "periphery" implies that secession is easier and potentially less costly. For convenience, without loss of generality, we order the $c_i(z)$'s in *z* so if $i > j$ then $c_i(z) > c_j(z)$. We also assume that for any *z*, $c_i(z) > c_j(z)$, then $c_i(z) > c_j(z) \forall z$. Finally, it is also useful to define the average cost function $\bar{c}(z) = \frac{1}{n}\sum_i c_i(z)$.

As discussed previously, one of the reasons that decentralized cooperation cannot occur is that observability of state shirking is imperfect. When states erect non-tariff trade barriers or enforce regulations arbitrarily, there are often disagreements on whether violations of federal agreemens have occurred. To capture the notion that monitoring of violations is imperfect, the second step in the stage game is that a non-strategic player reveals shirkers with probability *q(z),* where $q(z) \geq 0, q'(z) > 0, q''(z) \geq 0$. Here, the center's ability to monitor is a function of how strong its institutional power is: a weak center will not identify many shirkers, a strong center will identify more. Formally, 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 of the game is made by *C*, the central government, which simultaneously chooses how to enforce its power and how much of the contributions to return in the form of a "central" good. With respect to the latter, the good could either be a pure public good in which all states receive the same payoff, or one in which is excludable, but the center can provide more efficiently through scale. In the model, we formalize central good provision as *C* choosing a *payment vector* $\mathbf{x} = (x_1,x_2,...,x_N)$, which is the amount of payments made to each sub-unit. Note that again, these payments are indexed by *i*, meaning that the level of goods provision can differ by state. Indeed, as we note below, in the more restrictive case in which the good provided by the center is a pure public good, many of the intuitions gained from the model are strengthened.

We represent the gains from cooperation inherent in the federation as a production transformation technology, $\theta(n,z)$. The center's payments to the sub-units are modified by $\theta(n,z)$. We assume that a stronger center can better provide certain goods[5], so $\theta(n,z)$ is an increasing, concave function in *z.* In other words, $\theta(n,z)_z>0,\theta(n,z)_{zz}\leq 0$. To capture the notion of diminishing marginal returns to scale, we assume that $\theta(n,z)$ is a concave, increasing function of *n*, so $\theta(n,z)_n>0,\theta(n,z)_{nn}\leq 0$.

C also chooses a *punishment or extraction strategy* $\mathbf{m} = (m_1,m_2,...,m_N)$, which is a vector of indicators designating if an additional fee *f(z)* will be levied against each sub-unit *i*, where $f(z)\geq 0, f'(z)>0, f''(z)\geq 0$. These fines represent the power and resources granted to the center for enforcement of federal agreements. To simplify the analysis, we assume that the fines are "sufficiently high." In particular, we assume that for any *z*, $f(z)>c_N(z)$, so the fines are higher than the highest level of exit costs. The introduction of fines allows *C* to punish shirkers; but it may also use *f(z)* to extract rents from the states even when they do not shirk.

Finally, payoffs for the stage are determined and the stage ends. The payoffs of the actors are as follows. For a state *i,* its payoff in period *t* is:

$$u_{it} = \begin{cases} (1-s_{it})[\theta(n,z)x_{it} - f(z)m_{it}] - k_{it} - s_{it}c_i(z) & \text{if} \quad s_{is}=0 \quad \forall s<t \\ 0 & \text{otherwise} \end{cases}$$

[5] We use the term "central goods" to define the product of the center since our model allows for both public and non-public goods to be provided by the center. As long as the center can provide a good more efficiently (either because of its public nature *or* through scale effects) it will meet the criteria of our model. Thus we provide a general model in which the product of the center can be either provided in a discriminatory or a non-discriminatory fashion. This treatment of the central government's provision of goods being not purely public—in other words, including the possibility of 'local' discrimination—is similar to Tomassi (2000).

This formulation says the following. If a state is still in the federation in period *t*, it decides whether to remain in the federation (s_i=0). If so, it receives the amount granted to it by the center x_i enhanced by the central goods production parameter $\theta(n,z)$. If the center has penalized state (so m_i=1), the state must pay the center *f(z)*. Finally, state *i* decides whether to make a contribution to the center which costs it k_i. If a state *i* decides to exit (s_i=1), then it receives no contribution from the center, pays no fine *f(z)*, but must bear an exit cost $c_i(z)$. If a state has previously exited it earns zero in every period forward, so it obtains zero in period *t*.

The center has a utility function given by:

$$u_{Ct} = \sum_{i \in I_t} k_{it} - (1 - s_{it})[x_{it} - f(z)m_{it}] + s_{it}c_i(z)$$

where $I_t \equiv \{i | s_{is} = 0 \quad \forall s < t\}$. The center receives the sum of contributions from each state (k_i) less the transfer to each state from the benefits, x_i, net of any assessed fines $f(z)m_i$ applied to all states still in the federation. It also receives the exit costs from any seceding state, $c_i(z)$.[6,7]

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:

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

---

[6] We make three observations about the center's payoffs. First, the center collects fines levied against states. In many federations, this is how punishments are meted out. For example, in the European Union's Growth and Stability Pact, member states which are unable to meet deficit requirements must pay fines. Similarly, many federal policies in the United States reduce federal transfers to states that fail to comply with national rules. An alternative formulation that yields substantively similar results allows penalties to be a function of both *z* and *x*. Second, we also include benefits to the center when a state exits. The reason is that when the state enters a federal bargain, and carries with it exit costs, its bargaining power upon exit is reduced. In principle, the costs to the state from exiting may greater than the amount transferred to the center–indeed the center might actually also lose so this weight might be negative, but for now we ignore this complication. Our main purpose is to introduce *correlation* between rent extraction by the center and its ability to provide central goods and monitoring. Although we call these "exit costs," an alternative formulation would restate the propositions in terms of such a correlation and not exit costs. Finally, one might consider what happens when both the states *and* the center incur penalties or costs upon a state's exit. In this case, the equilibrium set is expanded; in other words the maximum amount required to keep the center in will increase. Substantively, this alters the comparative statics on exit costs but captures many of the same basic results we outline below. Finally, note that the fact that the center retains residual rents in each period implies an implicit intratemporal budget constraint.

[7] It is worth considering what central government's rents represent. These are of three sorts. The first and most obvious is corruption: personal enrichment by national political officials. A second source of rents is that the federal government may establish patronage systems and service to interest groups that gain it political support that can be used against the regions. Third, the center might collude with some group of states to extract rents and redistribute income from another group of states.

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

The sequence of moves in the *IG* is described in Figure 2.3. Here, the states confer to choose an institutional design. States make two choices. First, as before, they choose a constitution, embodying a set of rights and responsibilities of all of the members including the center. We model this as the states choosing a *punishment strategy cutpoint profile* **x**. 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, the states choose the parameter *z*, which is an argument in the exit cost functions $c_i(z)$, the fines that can be levied $f(z)$, and the center's production transformation function $\theta(n,z)$ in the repeated game. In the *IG* the states therefore choose how strong the center will be: increasing *z* at once increases the center's ability to provide central goods and enforce the agreed upon federal bargain, but also increases its ability to act opportunistically.

Finally, any state can choose not to participate, if the choices make the sub-national unit worse off than under no cooperative agreement. This structure represents *bottom-up federalism*: the states are designing rules to sustain cooperation.

3. Federalism as an Ongoing Concern

To solve this game, we use the equilibrium concept of *subgame perfection* which 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.* In this section, we assume that both the size, denoted by *n*, and institutions, denoted by *x* and *z,* of the federation are fixed, and solve for the equilibrium of the *RG*. Notably, *within* the *RG,* we cannot use backward induction, since the game has a positive probability of continuing at every point. Instead, we characterize classes of equilibria by positing the equilibrium strategies of the players and testing whether those strategies are optimal, given the other players' strategies.

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, and the center provides the equilibrium

level of central goods 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 1. *A player i plays a* **grim trigger strategy** *(GT) in each stage if:*

(i) on the equilibrium path, all states contribute, the center pays the equilibrium profile **x***, *and the center fines a state if and only if it is revealed a shirker;*

(ii) off the equilibrium path, if in the previous period, the center pays state i $x_i<x_i^*$ *or fines any state not revealed to be shirking, all states will exit*; *if any state exits, the center will set* **x**=0 *and* **m=1**.

The grim trigger strategy says that, if ever a player deviates from the cooperative equilibrium, all players irrevocably enter a defection stage. Under grim trigger, the players will cooperate only as long as all the other players have 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*, to establish 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 can be shown to hold for sufficiently long, finite punishments (as shown Bendor and Mookherjee 1987; see also Gibbons 1992). While analytically more convenient, *GT* yields substantively similar results to any other strategy in this class.[8]

In Proposition 1, we characterize the set of *GT* equilibria for the *RG* (all proofs appear in Appendix 1).

[8] It is important to note one proviso, however. While this approach to characterizing equilibria can be justified for our purposes here, it ignores an important consideration. By using grim, Nash-reversion strategies, this begs the question of why states cooperate in the punishments of others, even if the states are not harmed themselves. While this is certainly a central question to the design of federal institutions, we reserve a detailed discussion for other work. In order to provide some intuition, however, we examine the incentives for cooperation punishment in the appendix.

PROPOSITION 1. *Fix* δ,z,n. *If*

(A1) shirking constraint: $f(z)q(z)>1$, *and*

(A2) gains from federation: $\theta(n,z)>1$,

then there exist GT equilibria in which:

(i) states' cooperation threshold: $x_i^* \geq \frac{1-c_i(z)(1-\delta)}{\theta(n,z)} = x_i^{L*} \forall i$

(ii) center's cooperation threshold: $\frac{1}{n}\sum_i x_i^* \leq \delta - (f(z)+\delta\bar{c}(z))(1-\delta) = x^{U*}$

(iii) states contribute in every period,

(iv) and center fines only shirkers.

PROOF. In appendix.

Proposition 1 provides a number of insights into the ongoing dynamic between the center and the states and the conditions under which a federation can be sustained. Let us explain each of these conditions. First the shirking contraint condition $f(z)q(z)>1$ says that the expected fines from shirking must exceed the cost of contributing, so all states will contribute. Notice that because the parameters *f(z)* and *q(z)* are exogenous at this stage, either all states shirk or none do. This assumption thus defines a necessary condition for a stable federation: *the center must be given a strong enough to detect and punish potential shirkers*.[9] In addition, the condition implies that the constraint is more stringent as the function *q(z)* becomes smaller. The reason is that, the lower is *q(z)*, the higher must be *f(z)* in order that *f(z)q(z)* exceeds 1. This provides another prediction of the model: *grants of authority to the center are most likely to be sustainable in areas in which monitoring is relatively easy.*

Second, the gains from federation condition $\theta(n,z)>1$ implies that *there must be sufficient gains from exchange to motivate a stable federation*. The logic, however, is different from models of decentralized cooperation in which the benefit stream alone prevents individual states from shirking. In this case, the benefits have to be sufficiently large in order to gain a surplus that prevents the center from a one-time appropriation of all contributions.

Third, as long as conditions A1 and A2 are met, the federation is an equilibrium. Condition (i), the states' cooperation threshold, holds that every state must prefer the rents it receives from the center, $x_i\theta(n,z)$, to exiting. Not surprisingly, since x_i^{L*} is decreasing in $c_i(z)$, the minimum amount

[9] In a later section, we consider what happens when this ability to punish is at once correlated with the ability to obtain central benefits *and* an endogenous choice.

required to provide an incentive for a state to remain in the federation falls as the exit costs rise. Similarly, condition (ii), the center's cooperation threshold, states that the center must not be asked to return so much to the states, *in toto*, that it instead prefers to cheat while others are cooperating. The center's choice is between continuing to receive an ongoing payment from each state and taking all the contributions in the current period for itself, even though this implies losing all future payments. Maintaining the federation requires that center be sufficiently motivated, so it must pay out at most nx^{U*} in each period. Otherwise, it will appropriate all of the contributions for itself and fine all states, causing a breakdown in the federal structure.

Taken together, conditions (i) and (ii) mean that the set of equilibria depends on x^{U*} and x^{L*}, the upper and lower bounds on the average amount returned to the states by the center, where $\frac{1}{n}\sum_i x_i^{L*} = x^{L*}$. We illustrate these conditions in Figure 3.1. The states are arrayed continuously along the horizontal axis, by their level of exit costs. The heavy line shows the minimum level of contributions a state will be willing to accept before exiting x_i^{L*}. The line slopes downward since this quantity decreases as exit costs increase. The lower dashed line, then, represents the average exit level of these payments x^{L*}. The upper dashed line represents the maximum average amount that the center will be willing to return to the states x^{U*} and still participate; if the equilibrium payments are higher, then the center will destroy the federation by returning nothing. The difference between the two dashed lines represents the potential excess rent or surplus.

There are three possible relationships between x^{U*} and x^{L*}, each with an important implication for the types of federations that are sustainable. If $x^{U*} < x^{L*}$, then *no equilibrium exists*. In particular, there is no profile **x** which can at once keep all of the states in and provide the center with sufficient incentive not to deviate, to "take the money and run". In this case, federalism is impossible to sustain. In the knife-edge case, $x^{U*} = x^{L*}$, exactly one profile **x** can be sustained as an equilibrium: each subunit gets precisely its minimum amount x_i^{L*} in order to provide an incentive for it to stay in the federation, with the center keeping the remainder.

When $x^{U*} > x^{L*}$, there is a potential excess or "surplus" rent over the minimum amount required to sustain a federal agreement. This surplus rent must be divided between the players. In this case, a *multiplicity of equilibria* exists. Without more structure, it is not possible to say which equilibrium will prevail, a situation common in repeated games. Indeed, if the surplus rent, $S = n(x^{U*} - x^{L*})$, is positive, then any allocation of S that satisfies the states' cooperation threshold

is an equilibrium. Figure 3.2 illustrates this point.[10] It shows three possible equilibrium profiles **x**, all of which are consistent with the conditions in Proposition 1: (1) the allocation of the surplus equally among the states (i.e. $\frac{S}{N}$ to each of the state); (2) the allocation of all of the surplus *S* to a subset of the units; or (3) the allocation of all of the surplus to the center.

Fourth, if the costs of exiting are sufficiently high, the states have an incentive to remain in the federation, although the center does not pass on all of the rents to the sub-units. This indicates that *exit costs can shift economic and institutional authority from the states to the center*. Both the upper and lower bounds on x_i^* are decreasing as exit costs increase. As long as the center provides a positive value to the states, the states will remain in the federation. In sum, when the benefits are sufficiently large in relation to the exit costs, a stable federation can be sustained.

Fifth, using Proposition 1, it is possible to examine what factors affect the *size* of the set of equilibria with respect to the exogenous parameters.[11] With respect to the discount factor as the players value the future more, more profiles can be sustained in equilibrium (i.e. $\frac{\partial S}{\partial \delta}>0$, all proofs of these results are shown in the Appendix 1). This result is consistent with the folk theorem for repeated games, for as players value the future more, punishments in future rounds become more severe. The surplus or equilibrium set is also increasing in the productivity of the center (i.e. $\frac{\partial S}{\partial \theta}>0$). Here because there are more rents to distribute for a given level of contributions, there is more freedom (or surplus) which can meet the incentive constraints set by each of the actors. Alternatively, as the penalties which the center can impose increase, the size of the surplus decreases (i.e. $\frac{\partial S}{\partial f}<0$). The reason for this is that while *f* does not affect the lower bound required to keep a state in, it transfers rents to the center, pushing down the upper bound on payments necessary to keep the center cooperative. Thus, as *f* increases, the allowable surplus decreases. Finally, the size of the equilibrium set with respect to the average exit costs is ambiguous.[12] As shown in the appendix, increasing average exit costs decreases *both* the lower and upper bounds on x_i^*. If the lower bound falls faster than the upper bound, then the size of the surplus increases, otherwise it decreases. Thus, while

---

[10] This result extends in part from the fact that we analyze a set of equilibrium strategies in which all states are induced to punish the center even if the center transgresses or defects against only a subset of states. We take this approach for the reasons given above, allowing us to focus not on the multiplicity of deviations that might take place but instead on the minimal conditions necessary for cooperation. That said, our model is well suited to studying problems of coordination among states in punishments (see Bendor and Mookerjee 1987; Weingast 1997) which we reserve for later work. In Appendix 2, we provide an analysis of these issues.

[11] Here we mean how large is the surplus or excess rent and, therefore, the set of possible equilibria.

[12] Specifically, it is increasing iff $\delta>\frac{1}{\theta}$.

increasing exit costs shifts rents to the center, given that an equilibrium still exists, it can also make an equilibrium unobtainable.

Sixth, the *heterogeneity* in the states' cost functions means that the minimum level required to keep each state in the federation differs across states. For those states that have a large cost of exiting, the minimum the center will have to pay to induce them to continue in the federation is lower. This opens up the potential in some equilibria for the center to *price discriminate*.

Seventh, in terms of total social welfare, *all allocations are not equal*. In equilibrium, a typical state gets $\theta(n,z)x_i^* - 1$ and the center gets $n - \sum_i x_i^*$ in each period. Thus, if we define social welfare as the sum of benefits to all parties, the per-period total welfare is $\theta(n,z) - 1)\sum_i x_i^*$. Because $\theta(n,z) > 1$, this term is strictly positive in equilibrium. Further, social welfare is *increasing in* $\sum_i x_i^*$, the amount returned to the states. The reason is that the production technology benefit only accrues if *C* supplies central goods. Each unit which the center collects but does not return to the states represents an opportunity cost in public benefits forgone.

Finally, consider the shirking punishment strategies. Because the center gets utility from fines, in the one-shot game, the center will fine all states whether shirking or not. But in repeated play the states can counterbalance this incentive. If the center tries to extract too much through its enforcement technology, the states can credibly punish the center by exiting. If the benefits from ongoing cooperation with the states are high enough, the center will not extract "inappropriate fines."

4. Endogenous Institutions

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 *RG*. 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, provide part of the way out of this quandary. In bottom-up federalism, the states have a say in the design of federal institutions and hence in federal performance.

In this section, we use the results from the previous section to solve the institutional game, *IG*. States erecting a bottom-up federalism will "look down the tree" at the *RG* and will choose institutions that are efficient. In the *IG*, the states do three things. First, they choose an equilibrium profile of triggers **x** that determines the minimum level of central returns to each state to avoid triggering a punishment phase. The division of potential surplus rents is unidentified in the model specified thus far. In order to pin these down, we use a simple Nash bargaining framework in which each state has a certain amount of *pre-play bargaining power*, in order to determine the division of rents. Thus, we designate the vector $\alpha=(\alpha_1,\alpha_2,...,\alpha_N)$ as a vector of individual bargaining weights, where $\alpha_i \geq 0 \forall i$ and $\sum_i \alpha_i = 1$. This allows us to examine very general divisions of the rents between the states.

Second, states collectively choose the level of institutional authority z to grant the center.[13] This choice reflects a fundamental trade-off in federalism. Assuming that the states can motivate the center to return a significant part of the payments to themselves, then a higher z means a higher θ, yielding larger benefits per unit for the states. Yet a higher z also increases the exit costs and the potential fines, meaning that the center can extract more rents from the states. Third, just as states have an option to exit at every stage of the *RG*, in the *IG,* states have the option of *not entering* the *RG.*

Because we examine here an exogenously determined set of states in the ongoing federation, we use the following solution concept. (We endogenize state participation in the next section.) We characterize the set of equilibria such that all states must want to participate, given a *cooperative GT* equilibrium exists to the *RG.* In other words, the choice of the equilibrium must be Pareto efficient among the states.

Using this solution concept, we have the following result:

PROPOSITION 2. *Fix n and assume there exists a z such that:*

(A1) positive surplus: $\delta - (f(z)+\bar{c}(z))(1-\delta) \geq \frac{1}{\theta(n,z)}$

(A2) no shirking: $f(z)q(z)>1$

Then a GT equilibrium exists that has the following IG equilibrium properties:

(i) optimal trigger: $x_i^* = (1-\alpha_i n)\frac{1}{\theta(z*)} + \alpha_i n(\delta - (f(z*)+\bar{c}(z*))(1-\delta))$

[13] This concept means that, conditional on the existence of an equilibrium to the *RG*, we characterize the *core* of *(***x**, *z).*

(ii) optimal central power: z^* *solves* $\frac{\theta_z}{\theta} = \frac{(f_z + \delta\bar{c}_z)(1-\delta)}{\delta - (f+\delta\bar{c})(1-\delta)}$ *and has a unique solution*

(iii) all states participate.[14]

PROOF. In appendix.

Before turning to the implications of the analysis, consider the conditions (A1) and (A2). Condition (A1) simply states that S is positive, so an equilibrium can exist. Condition (A2) says that there is some level of institutional power such that shirking will not overwhelm the federation. Together, these conditions accomplish two things. First, they guarantee that an equilibrium to the *RG* exists, on which basis it is possible to examine the institutional choices made in a bottom-up federation. In addition, they guarantee that the solution to institutional design problem will be an interior one.

Proposition 2 yields a series of important implications about an equilibrium federation. First, in a federation, a *constitution may act as a focal point that defines the limits on central authority*. A set of decentralized states face a coordination problem: if the definition of central transgression is unarticulated, then states may fail to coordinate on their punishments of the center, ultimately causing the federation to unravel. The choice of a set of cutpoints that trigger punishments, **x**, can overcome this coordination problem. When erected prior to playing the federalism game, a constitution can serve as a focal, coordinating device by determining precisely what constitutes central encroachments (see Chen and Ordeshook 1994; Hardin 1989; and Weingast 1997).

Second, all states have one interest in common: they want to maximize the size of the surplus to be distributed among themselves. *States will therefore choose a punishment strategy,* ***x****, that provides the center with the minimal level of rents in order for it to cooperate*. This implies that they capture the remainder of the rents for themselves collectively; that is, $\sum_i x_i^* = nx^{u*}$. The opportunity to establish focal strategies gives an institutional advantage to the states over the center. 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 1997).

Third, making participation endogenous to the federal bargain *increases the states' lower bound of acceptance of a federal bargain* from the earlier game. Whereas before, high exit cost

[14] Note, in condition (ii) we use the convention of subscripts of endogenous variables to indicate the first derivative with respect to that variable. We also suppress the arguments of the functions in condition (ii) for expositional simplicity.

states would *continue* in a federation even if their payoffs were less than their contribution, here states will not *enter* the federation if the equilibrium payoffs are not at least as high as they could obtain in the absence of the federation. This raises the lower bound on each state's payoffs from x_i^L to $\frac{1}{\theta(z*)}$. To see this, note that a state outside the federation earns zero in each round. At a minimum, therefore, a state will enter the federation only if its equilibrium stage game payoff $\theta(z*)x_i^* - 1 \geq 0$. Figure 4.1 illustrates this result. Fixing z^* according to (ii) in Proposition 2, the average payoff to the states will be the minimum required to provide the center with the incentive to stick to the federal bargain, denoted by $x^{U*} = \delta - (f(z) + \bar{c}(z))(1-\delta)$. Without a participation constraint, the minimum any single state can receive in equilibrium is $x_i^{L*} = \frac{1 - \bar{c}(z)(1-\delta)}{\theta(z)}$, represented in the figure by the heavy solid line. With the participation constraint, however, every state must receive at least $\frac{1}{\theta(z*)}$ represented by the heavy dashed line.

This contrast highlights an important feature of federal institution building. *Ex post* there can be significant differences between states vulnerability to rent extraction, due in our model primarily to heterogenous exit costs. Adding a participation constraint allows states with higher exit costs to reduce the potential for ex post opportunism through ex ante bargaining over the institutions. Unlike in the *RG,* in which participation was fixed, each state's minimum return here is identical.

Further, the result means that a *federation is even harder to sustain* than implied by the results of the previous section.[15] Before, as long as the minimum required to meet the center's cooperation incentive averaged the same as the lower bounds on the ex post requirement for the states, cooperation could be sustained as an equilibrium. Adding institutional choice means that x^{U*} must be much higher: strictly greater than the maximum x_i^{L*}.[16]

Fourth, result (ii) above *highlights the central tradeoff in a federal system*. The choice of z is the result of a maximization problem for the states. States have a common interest in a strong center: as the center becomes stronger (reflected on the left hand side of the equality), the shirking problem is more easily solved *and* the center's ability to provide central goods increases. Yet a strong center is also able to appropriate a greater portion of the transfers. The solution to this problem is to equate these two at the margin: set z so that the marginal benefits from the center's

---

[15] By "harder" we mean in the sense that the parameter space over which a cooperative outcome can be maintained is smaller.

[16] To see this simply note that $\frac{1}{\theta(z)} > \frac{1 - c_i(z)(1-\delta)}{\theta(z)} \forall i$ since $c_i(z) > 0 \forall i$.

prevention of shirking and central goods provision equal the marginal costs of increased rent extraction.

Notice that in a bottom-up federations, the choice of z does not involve a distributive conflict: all the states have a common incentive to maximize the surplus in our model, each garnering a fixed proportion. Further, the assumptions of Proposition 2 imply that the solution z^* is a unique optimum. Put another way, the parameters δ,c,f, and q, imply a unique set of institutions for each federation.

Finally, the model yields predictions about the nature of the central institutional authority as a function of the parameters and functions in the model. By implicitly differentiating result (ii) in Proposition 2, we have that *z^* is decreasing in average exit costs* (all proofs in Appendix 1). This leads to a significant prediction to the model: in bottom-up federations in which the *ex post* costs of exit are high, we should expect to see weaker institutions, a lower provision of central goods, and less social welfare. Similarly, just as average exit costs shift rents toward the center, so do fines. This again creates a disincentive, all other things equal, for the states to cede more institutional authority to the center. In other words, *z^* is decreasing in the ability of the center to impose penalties.* Finally, *z^* is decreasing in the productivity of the center, and therefore decreasing in n.* Here the logic is slightly different even though the outcome is the same: because the center can better produce central goods, there is no need to cede as much control to the center all things equal.

5. Equilibrium Federations

In the previous section, the structure of the federation was fixed and participation was considered only from the perspective of the state itself. Each state must have had an incentive to participate or an equilibrium could not be sustained. In practice, the choice of participation goes both ways: not only must an individual state opt in to a federalism, but also the other states must consider whether or not to include the marginal state. In this section, we adapt our model to analyze federal exclusivity. Here the conditions for an equilibrium are more stringent: all parties must prefer each included member to be in (including the member in question) or the federation is not an equilibrium. In this sense, by endogenizing the participants–allowing the size and character of the federalism to vary–in addition to defining equilibrium institutions and ongoing actions, we define, in terms of our model, the characteristics of "equilibrium federations."

To consider equilibrium federations in the spirit outlined above, we analyze the set of possible federations given the exogenous characteristics of the constituent units. To analyze this problem, we consider the conditions under which a set of *n-1* states will choose to include the *n*th possible member in the federation. The solution concept we employ is an incremental variant of coalition-proofness. In this case, if every state in the set *1,...,n-1* is better off from the inclusion of state *n*, and state *n* is also better off, then the *n-federation* dominates the *n-1* federation and is said to be an equilibrium federation. In addition, as the analysis in the previous section showed, we also need to define the division of the surplus in the old and new federation. In both cases, we assume that the surplus will be divided according to the same weights as before, with the new member receiving none of the surplus (in other words $\alpha_N=0$). This will allow us to find the cases in which the inclusion of the new member is most likely.

Using this concept we can state the following proposition:

PROPOSITION 3. *Let* $\bar{c}^{n-1}$ *and* z^{n-1} *be the average exit cost function and equilibrium institutional strength of the n-1 federation of states. Define* $\bar{c}^{n}$ *and* z^{n} *be defined similarly for the n federation. Then the n federation dominates the n-1 federation iff:*

$$n\theta(n,z^n)[\delta-(f(z^n)+\bar{c}^n(z^n))(1-\delta)]\geq(n-1)\theta(n-1,z^{n-1})[\delta-(f(z^{n-1})+\bar{c}^{n-1}(z^{n-1}))(1-\delta)] \qquad (*)$$

PROOF. The result follows directly from *Proposition 2*.

The expression (*) simply states that the *n* federation dominates the *n-1* federation if the surplus rent is greater under the larger federation. It is useful to determine when this condition will hold. On the one hand, there are scale benefits to the larger federation. Note that our model highlights two types of scale benefits. The first is the direct benefit that the surplus itself, all things equal, will be larger with more states. The second benefit is the scale advantage in productivity. Since the production technology exhibits increasing, albeit diminishing marginal returns to scale, increasing the size of the federation will improve the overall productivity of the federation. So even in this case when all the incremental benefits accrue to the existing *n-1* members–in other words, that the new member is made indifferent by joining–it might seem that the benefits that accrue are always positive.

In fact, this might not be the case. The reason is that scale benefits relative to the *n-1* federation will be traded off against a potential reduction in the *average* level of returns from a stronger center. The left hand side of the expression captures this potential relative loss. What factor is the critical determinant? If the marginal state has high exit costs (raising the average level of these costs), two adjustments, depending on the parameter values, can take place: first, the new federal institutions will have weaker central authority since z is decreasing in exit costs and second, the higher exit costs will increase center's incentive to deviate.

We illustrate this tradeoff in Figure 5.1. The two heavy lines provide an illustration of the analysis in Section 4: in the *n-1* federation, the states will maximize the difference between the costs of increasing central authority (as represented by rent diversion to the center through a lower x^{U*}) and the benefits centralization create from scale which is a function of z as well. The result will be a choice of z that maximizes the difference between these two, creating a large surplus, at z^*. An entering high exit cost state has two effects: first it shifts up the "cost" curve which determines in part the transfers to the center; second it also moves the benefits curve through an increase in n. States will optimize the institutions z in the same way as before, possibly shifting z down to z^{**}. Finally, given all of these effects, if the difference at the new optimal level of z is *smaller*, then the *n-1* federation will refuse entry to the new state. Otherwise, the federation will expand.

This result raises a number of important implications for the nature of political agglomeration and dispersion. Most importantly, the trade-off here is similar to other models (e.g. Alesina and Spolaore 1997, Alesina and Wacziarg 2000, Bolton and Roland 1997, Tiebout 1956): as with those models, there is a balance struck between the benefits of scale and the costs of less flexibility. What differs here, however, is that we focus on a particular type of loss: *a deviation from optimal institutions*. In our model, the growth in federalism (or lack of it) is driven by the nature of *ex post* opportunism by the center.

Further, Proposition 3 also yields predictions about the nature of incremental federal agglomeration, either by the increase in portfolio jurisdiction ceded to the center, or by the inclusion of new states. The model generates the prediction that *as exit costs rise, federal institutions will be weaker*. In cases such as the European Union, for example, this would mean that one of two things were happening with expansion: either the ability of the center to extract

rents (e.g. *f(z)* and *c(z)*) were declining, or the authority of the center was reduced (*z*) as the federation grew.

6. Three Cases of the Dilemmas in Action

The previous models provide a number of important predictions about the relationship between state characteristics, the opportunities for collective benefits to be captured, and the stability and character of federal institutions. Although a detailed empirical analysis is beyond the scope of this paper, a few empirical illustrations provide both an understanding of the motivation behind the assumptions of the model, illustration of the mechanisms at work, and the plausibility of the predictions derived from them.

In this section we focus on one of the main predictions of the model. According to the results, the ability to overcome the two dilemmas depends on how sharp the tradeoff is between moral hazard by the states and rent-seeking by the center. If the tradeoff is too sharp, the federation will degenerate into one which is completely peripheralized or completely centralized, in Riker's terms.

Each of the three examples–the creation of the American federal Constitution, the American Nullification crisis, and modern Chinese federalism–illustrate how this tradeoff can both doom a federation, and how relaxation of the tradeoff can lead to a self-enforcing, cooperative federation. Each case contains two periods. In the early stage, the moral hazard problem by the states could not be solved because the technology to monitor the states would require too great an opportunity for the center to use those powers to obtain rents. Indeed, the resolution required institutional innovation that relaxed the tradeoff: only by limiting the abuse of enforcement powers could the center be credibly trusted to provide monitoring and public goods without the danger of aggrandizement.

6.1 American Federalism: From Articles of Confederation to Constitution

Nearly all the major turning points in American history can be studied from the perspective of federalism. Federalism is central to the revolutionary crisis, the debates over the Constitution, the Civil War, Reconstruction and its end, the New Deal, and the rise of the regulatory state in the 1960s and 1970s.

The principal criticism of the Articles of Confederation by Federalist leaders was that the national government had insufficient institutional power to supply critical central goods, primarily defense against British and European security threats, but also the maintenance of public economic structures, such as a common market and a common, stable currency. One of the core debates between the Federalists and Anti-Federalists concerned how to provide these goods.[17] The Federalists believed that the national government should be granted strong taxation powers in order to have resources to achieve these ends (Kaplanoff 1991, Morgan 1977, ch 9). Some Anti-Federalists admitted a concern about the under supply of central 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.[18]

Under the Articles of Confederation, the Anti-Federalists' political power allowed them to maintain the balance in their favor. For example, although Congress passed defense bills, it could not raise money to finance these measures. Instead the national government depended on the states to raise taxes to finance national legislation. Because the center had insufficient enforcement powers, many states simply refused to contribute (Kaplanoff 1991, Middlekauff 1982).[19] Similarly, control of currency was also impossible. Rhode Island, for example, refused to discontinue its practice of "over supplying" and thus devaluing currency, hindering the center's ability 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. All three national public goods — adequate security, a sound currency, and the common market — suffered because of state free-riding in the face of the common pool problem.

---

[17] As Hamilton outlined in *Federalist No. 23*: "The principal purposes to be answered by union are these–the common defense of members; the preservation of the public peace, as well against internal convulsions as external attacks; the regulation of commerce with other nations and between the States; the superintendence of our intercourse, political and commercial, with foreign countries." (Hamilton, Madison and Jay 1961: 153)

[18] As Rakove (1996,146) emphasizes, the Anti-Federalists "favored lines of attack evoked customary Whiggish fears of concentrated power and the specter of a potent central authority absorbing the residual powers of the state governments.

[19] It is worth recalling that the Federalists opened their famous debates with an extended discussion of the problems of national defense under the Articles (Hamilton, Madison and Jay 1961, *Federalist Nos. 2-5*). Although these are not nearly as widely cited as those focusing on institutions, it is no accident that the Federalists opened with this topic (see Riker 1987).

The veto structure of decisionmaking under the Articles — in which single states could block passage of national programs — meant that the national government had insufficient power to provide goods and enforce contributions. One of the main problems with the Articles was that they did not clearly define the limits of federal authority. The Federalists' proposal to grant the national government additional taxation power failed to create limits on how far this power could be taken. Our model suggests that the Anti-Federalists rightly feared predation with a stronger national government under the Articles. They therefore blocked Federalist initiatives to increase national power. The result was an ineffectual federation from 1781 to 1789 (Middlekauff 1982, ch 23; Morgan 1977, ch 9).

The genius of the Federalists in creating the new Constitution was in the way they resolved this dilemma through institutional rules. The Constitution first granted the national government sufficient power to provide the critical national central goods of national defense, common markets and common currency. Second, it created limits on the national government, thus constraining the national government's use of the enforcement powers granted to it by the states.

Limits on the national government took several forms. First, the Constitution contained a series of 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 made it hard for extremists to take control of the national apparatus, so that "ambition would check ambition" (see Madison, *Federalist No. 46*, 1961: 294-300). This system was reinforced by having an institution, the Senate, to represent the each states directly. Similarly, the Supreme Court was established with the authority to enforce these rules.[20]

Second, the debates during the Revolutionary crisis and over the Constitution helped forge a consensus about how to limit the national government.[21] If the national government over-stepped these limits, states would threaten to secede.[22] In *Federalist No. 46*, Madison (1961: 298) made explicit reference to the coordination the constitution would provide to prevent central aggrandizement. "But ambitious encroachments of the federal government on the authority of the

---

[20]As Bednar, Eskridge, and Ferejohn (2001) show, the Court was better at policing the states than the national government.

[21]Rakove, Rutten, and Weingast (2000) make this point; see also Rakove (1996) and Wood (1969).

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

State governments would not excite the opposition of a single State, or a few States only," he stated. "They would be signals of general alarm. Every government would espouse the common cause. A correspondence would be opened. Plans of resistance would be concerted. One spirit would animate and conduct the whole."

To illustrate the use of these "trigger strategies", consider the controversies raised under the Federalist President, John Adams, and his Secretary of the Treasury, Alexander Hamilton. Federalists sought to expand the national government powers in controversial ways, notably to promote economic development (Elkins and McKitrick 1993, ch 7; Lenner, 2001, ch**). And yet, in the late years of the eighteenth century, the Federalist's popularity waned. The Adams administration reacted with, among other things, the Alien and Sedition Acts of 1798. These acts in part 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. Many Federalist supporters switched sides to support the opposition, allowing Thomas Jefferson to become president in the election of 1800. Indeed, the Federalist's behavior not only helped foster the development of an opposition party, but to spring-board it into power for twenty years (Wood 1992). 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 1997).

Our model provides considerable insight into the American founding era. The national government was too weak under the Articles. Yet Federalists' attempts to strengthen the national government's power failed due to Antifederalist rational fears about abuses from the center. The Federalists resolved this problem with the new Constitution that simultaneously increased the national government's powers while creating a set of explicit checks, balances, and explicit limits on the that government.

6.2 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 in 1800.[23] As with the controversy over creating the American Constitution, that surrounding nullification focused on the appropriate bounds, both upper and lower, on national government power.

To understand the genesis of the nullification crisis, we begin with the crisis over the admission of Missouri from 1819 to 1820, which demonstrated that the previous consensus had collapsed (Syndor 1948,132-33). The crisis began when Southerners, then the more dominant section, sought to admit an additional slave state, Missouri, without any free state to balance it.[24] Many Northerners feared that the loss of balance would allow Southerners to dominate the national government at their expense. Northerners feared Southern use of national power to extract rents from them.

Northerners reacted by attacking Southerners where they were most vulnerable: slavery. They amended the legislation to admit Missouri in the House of Representatives, where they held a majority. The amendments prohibited the further importation of slaves and provided for a gradual emancipation of slaves already resident in Missouri. These provisions failed the Senate, where the South held a veto, and a crisis ensued.

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. By the end of the 1820s, John C. Calhoun, an ardent nationalist in the early years of the second decade of the nineteenth century, had become the major proponent of states' rights.

In short, both sections had reason to fear the other's control of the national government. Without a veto over national policymaking, each was vulnerable to encroachment by the other.

The Missouri Compromise of 1820 resolved the crisis with three components: immediately, it balanced the admission of Missouri by carving off the northern counties of Massachusetts 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 (Meinig 1992, Weingast 1998).

---

[23]This section reports on research in collaboration with Douglas Grob.

[24]Moore (1953) provides the standard history of this controversy; see also Fehrenbacher (1979, 1980) and Freehling (1990).

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.

Establishing sectional balance was a minimal condition to provide national stability. As each section feared the other's control over the national government, a veto over national legislation allowed them to constrain the national government's powers and hence potential abuses.

Despite the balance rule, many radical Southerners still feared the designs of the North.[25] During the first Jackson administration, radical Southerners, such as Calhoun in South Carolina, proposed a new check on national authority known as the nullification a doctrine. Using a variant of proposals offered by Jefferson and Madison during the Adams's administration in response to the Alien and Sedition Acts, nullifiers argued that a state could interpret and defend the Constitution on its own, affording it the power to "nullify" or set aside national legislation within their borders. Calhoun further claimed that the Constitution was not a "forever pact," but a compact among sovereign states that could exit. The nullification doctrine meant that states could pick and choose which national legislation they 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. Had nullification been upheld, it would have eliminated the national government's ability to impose and police standards and hence to police shirking. The result would have been free-riding and breakdown of American federalism. Second, nullification would have drastically facilitated a states' exit: indeed, its titular purpose was to allow costless exit.

Ultimately, Jackson and his political advisor and organizational genius, Martin Van Buren, defeated the nullifiers. Jackson helped forge a near national consensus over a new approach to states' rights (Freehling 1966). The new approach held that the national government had virtually no role in regulating the economy, except through taxation to provide enumerated

---

[25] Standard sources on nullification include: Ellis (1987) and Freehling (1966); see also Freehling (1990, chs 14-15) and Knupfer (1991, ch3).

central 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.

Jackson's veto in the controversy over renewing the charter of the Second Bank of the United States illustrates this claim. Jackson's famous veto message went well beyond the specifics of the Bank controversy to create a defining moment for the Jacksonian Democrats. This event is important for two reasons. First, Jackson had to use his veto because sufficient Jacksonians joined the opposition to pass the renewal. Second, the veto message articulated the new approach of states' rights in a way that limits the power of the national government. The approach sought not only to allow states the freedom over their property and institutions, but to deny the federal government any powers over economy that might be used as a precedent to grant that government power over slavery. Jacksonians had to oppose the Bank because they denied the federal government the authority to intervene in the economy. Third, the new approach to states' rights helped forge a new consensus about the limits on national authority and hence trigger strategies to engage against potential encroachment by the national government.

The new consensus over states' rights helped create new, self-enforcing limits on the national government. It also 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 6.1 reveals, Democrats held united control of the national government in 8 of the 16 Congresses 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. Try as they might to develop support for greater national promotion of the economy, Whigs could never muster sufficient electoral support (Holt 1999). Second, as long as these doctrines maintained the Democrats' dominant electoral position, they 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.

Our model helps interpret the nullification controversy. Many southerners became concerned in the 1820s about the security of their property and institutions within the Union. They proposed nullification as a new tool to protect themselves. Yet this tool had many larger consequences, lowering the institutional authority of the center, creating new common pool problems, and generally returning the country to the situation under the Articles. Jackson resolved this crisis by creating new and sustainable limits on the national government. This helped him isolate the radicals and forge a majority coalition that could implement his solution for a generation. This controversy also illustrates our point about exit costs, for Jackson was willing to use force to keep South Carolina in the Union.

6.3 Modern China

Mao's death in 1976 left China in disarray. The cultural revolution had been an economic and political disaster. Further, Mao's 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. 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 (Oi 1993, 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 (Oksenberg and Tong 1991). 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 for locally generated tax revenue. Because they could keep most or all tax revenue beyond a certain amount, lower governments had strong incentives to foster local economic prosperity. Economic growth would benefit local citizens and local governments, not just 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 (Montinola, Qian, and Weingast 1995).

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) (see Bahl and Wallich 1992 and Wong 1991). 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 1988); not to subsidize, creating the so-called soft budget constraint (Dewatripont and Maskin 1995; generally, see Riordan and Aghion 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, "storing wealth in enterprises".

China's policies for creating economic and political reform created a new system of 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 (since 1978). 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 backed down. China's new system of federalism survived its biggest challenge. The threat of non-cooperation by the provinces proved central to policing the center's willingness to adhere to the federation's rules.

From the beginning, the central government represented the principal impediment to fostering a market economy. Under the Chinese socialist system, the central government's institutions were geared toward command and control, not the market. Hence, per our model, the first steps in market reform were to reduce the reach of government. The new fiscal system proved both a significant constraint on the center and a new focal point around which the provinces could coordinate to police the center's encroachment on their new rights.

7. Conclusions

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

This paper develops a model of self-enforcing federalism, showing how stable federations solve the two fundamental dilemmas of federalism. Our models yields a series of results. First, for a federation to overcome the shirking problem, the center must have sufficient monitoring resources and penalizing capacity to punish shirkers. Second, to police the center's tendency to overawe the states, states must *coordinate* on punishment strategies, perhaps chosen at the constitutional or design stage of a federation. Appropriately designed punishment strategies limit the center's ability to extract resources from the states, increase the provision of public goods, and result in higher public welfare. Third, *exit costs* shifts rents to the center. As a state's cost of exiting increases, its threat to exit becomes less credible. This increases the bargaining power of the center against the state, and shifts some of the rents to the center. Fourth, 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. 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. An inappropriate level of institutional power granted the center is 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 federations (Elazar 1987,192-97). 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 a federation.

Our approach also suggests an important difference between top-down and bottom-up federations. As Stepan (1998) emphasizes, top-down federalism includes much of the recent

trend toward decentralization. Although space does not allow an extended discussion, our model yields several important results about top-down federations. A federation designed by the center is likely to leave the center with a greater share of the rents than a bottom-up federation. The reason concerns who holds agenda power. In bottom-up federalism, the constituent states design the federation and will attempt to choose institutions that capture the rents for themselves. In top-down federalism, the center controls the design and will bias institutions in favor of its interests.

This perspective on top-down federalism yields a comparative statics result, which applies to the recent literature on the break up of nations (Alesina and Spolaore 1997, Alesina, Spolaore, and Wacziarg 2000). Consider a top-down federation in which the center has designed the institutions to maximize its share of the rent. This implies that the marginal state is indifferent between remaining or exiting the federation. Next, suppose that exit costs fall, so that the marginal state now has an incentive to exit. In response, the center is likely to adjust the costs and benefits of federalism so that the marginal state will remain in the federation.

Alesina et al. study the growth of international trade, suggesting that, by providing a substitute for the scale benefits of a large country, growing international trade lowers exit costs for regions in federations. They predict that this will lead to the break up of nations. We disagree, observing that Alesina et. al. ignore the endogenous reaction of the center. In response to falling exit costs, the center is likely to increase the benefits to marginal regions, for example, by increasing authority to the states. Thus our prediction is that, in response to growing international trade and lower costs of exit, heterogeneous countries should decentralize.

In Section 6, we demonstrated both the strengths and weaknesses of our theoretical results through the examination of a series of cases: the failure of federalism in the United States under the Articles of Confederation and how the Constitution resolved the twin dilemmas of federalism; the reappearance of these problems during the nullification crisis; growing decentralization in China; and the failure to solve these problems in other large federations, such as Argentina, Mexico, and Russia. 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. Per our predictions, the successful cases — the United States and to some extent China — resolve the twin dilemmas in accord with our model: creating a clear delineation central power while granting the center the power to police shirkers. The failed federations — Argentina, Mexico, and Russia — have failed to counterbalance central authority.

Our paper contributes to the growing literature on "equilibrium institutions" (Calvert 1992 and Gibbons and Rutten 1996). This approach holds that, to be sustained, all features of representative government must be self-enforcing in the sense that political officials have incentives to abide by them. This logic includes sustaining political institutions — such as, elections, separation of powers, and federalism — and various rights — such as the right to hold property, to religious freedom, and to form free associations. Our approach to federalism demonstrates the power of such a perspective. Using the formal tools of rational choice institutionalism, we focus attention on the specific trade-offs and requirements of stable federal institutional arrangements. To survive, the federal institutions must be self-enforcing for political officials at all levels of government.

More generally, for students of constitutions and democratic institutions, we use the case of federalism to demonstrate how to study a neglected aspect of constitutions. The vast majority of the literature examining constitutional institutions takes these rules as exogenous. In contrast, the new literature on equilibrium institutions takes these institutions as endogenous and seeks to explain the factors underpinning their survival. By taking the approach that constitutions should be studied as self-enforcing equilibria, we have demonstrated not only the force of such documents but also their rationales.

APPENDIX 1: PROOFS OF PROPOSITIONS STATED IN TEXT

Proof of Proposition 1.Consider first a typical state *i*'s cooperative strategy in equilibrium. Consider first the payoff to shirking versus cooperating. The payoff it will earn for shirking for one period will be $\theta(n,z)x_i - 1$. Its payoff for contributing will be $\theta(n,z)x_i - f(z)q(z)$. Solving for these two conditions implies that a player will contribute over shirk iff $f(z)q(z)>1$. Now consider when it will contribute versus exit. If it exits it's payoff will be $-c_i(z)$. If it contributes its expected payoff will be $\sum_{t=0}^{\infty}\delta^t(\theta(n,z)x_i - 1) = \frac{(\theta(n,z)x_i - 1)}{1-\delta}$. Thus, a player will cooperate rather than exit iff $x_i \geq \frac{1 - c_i(z)(1-\delta)}{\theta(n,z)}$. Now consider the equilibrium strategy of the center. It is straightforward to show that given the equilibrium strategy of the states, the center's dominant strategy is to play is $x_i = 0 \forall i$ and $m_i = 1 \forall i$. Thus the payoff to deviating for the center is $\sum_{t=0}^{\infty}\delta^t\sum_{i=1}^{n} f(z) + \delta c_i(z) + 1 = n(1 + f(z) + \bar{c}(z))$. Its expected payoff to not deviating is $\sum_{t=0}^{\infty}\delta^t\sum_{i=1}^{n}(1 - x_i) = \frac{n}{1-\delta}$. This in turn implies that the center will stay on the equilibrium path if $\sum_{i=1}^{n} x_i \leq \delta - (1-\delta)(f(z) + \bar{c})$. To determine enforcement off the equilibrium path, consider first the Nash equilibrium in the stage game. As noted the center's dominant strategy is $x_i = 0 \forall i$ and $m_i = 1 \forall i$. Note also that given the center's optimal strategy, the states will always prefer shirking to contributing, since $-(1 - f(z)) < -f(z)$. Now consider the state's choice of exiting versus shirking. A state will prefer to exit over shirk in the stage game iff $-c_i(z) > -f(z) \Rightarrow c_i(z) \leq f(z)$, which is true by assumption. Thus, since the off-path equilibrium strategies are reversion to the Nash equilibrium, enforcement is subgame perfect.

Proof of comparative statics in the RG. Note first that $x^{U*} = \delta - (f(z) + \delta\bar{c}(z))(1-\delta)$, $x^{L*} = \frac{1 - c(z)(1-\delta)}{\theta}$ and $S = x^{U*} - x^{L*}$. This implies the following: *(i)* $\frac{\partial x^{U*}}{\partial \bar{c}} = -\delta(1-\delta) < 0$ and $\frac{\partial x^{L*}}{\partial \bar{c}} = \frac{(1-\delta)}{\theta} < 0$; *(ii)* $\frac{\partial S}{\partial \bar{c}} = -(1-\delta) < 0$. *(iii)* $\frac{\partial S}{\partial \theta} = \frac{1}{\theta^2}[\delta - (f + \delta\bar{c})(1-\delta)] - \frac{1}{\theta^2}[\delta - (f + \delta\bar{c})(1-\delta) - \theta(1 - \bar{c}(1-\delta)]$. Substituting the expressions for x^{U*} and x^{L*}, this simplifies to $\frac{\partial S}{\partial \theta} = \frac{(\theta - 1)x^{U*} + \theta x^{L*}}{\theta^2} > 0$ since $\theta > 1$; *(iv)* $\frac{\partial S}{\partial \bar{c}} = \frac{(1-\delta)(1-\theta\delta)}{\theta} \rightarrow \frac{\partial S}{\partial \bar{c}} > 0 \Leftrightarrow \delta > \frac{1}{\theta}$.

Proof of Proposition 2.Note first that assumptions (A1) and (A2) guarantee that an equilibrium to the *RG* exists. Now consider a typical state *i*'s participation constraint. A state will participate iff her equilibrium stage payoff is greater than zero which implies $\theta(n,z,)x_i^* - 1 \geq 0 \Rightarrow x_i^* \geq \frac{1}{\theta(n,z,)}$. This implies that each state will receive $\frac{1}{\theta(n,z,)} + \alpha_i S$. If we solve for each state *i*'s preference for *S*, we have $\max_S \frac{1}{\theta(n,z,)} + \alpha_i S$ subject to $S \geq 0$ which implies $x^{U*} = \delta - (f(z) + \delta\bar{c}(z))(1-\delta)$. Solving for x_i^*, we have $x_i^* = \frac{1}{\theta(n,z)} + \alpha_i n[\delta - (f(z) + \delta\bar{c}(z))(1-\delta) - \frac{1}{\theta(n,z)}]$ which is part (i) of the proposition. To find the optimal *z* for a given state *i*, we must maximize the sum of the discounted equilibrum payoff, which implies a state's optimal *z* can be obtained by maximizing the sum of its stage payoff. Taking

$$\max_z(\theta(n,z)x_i^* - 1) = \max_z[\theta(n,z,)(\frac{1}{\theta(n,z)} + \alpha_i n[\delta - (f(z) + \delta\bar{c}(z))(1-\delta) - \frac{1}{\theta(n,z)}) - 1] \qquad \text{(A1)}$$

we have the condition

$$\alpha_i n(\theta_z(n,z)(\delta-(f(z)+\delta\bar{c}(z))(1-\delta)-\theta(n,z)(f_z(z)+\delta\bar{c}_z(z))(1-\delta))=0 \qquad \text{(A2)}$$

which implies that for player i, z_i^* solves

$$\frac{\theta_z}{\theta}=\frac{(f_z+\bar{c}_z)(1-\delta)}{\delta-(f+\bar{c})(1-\delta)}. \qquad \text{(A3)}$$

The second order condition of (A1) is

$$\theta_{zz}(\delta-(f+\bar{c})(1-\delta))-2\theta_z(f_z+\bar{c}_z)(1-\delta)-\theta(f_{zz}+\delta\bar{c}_{zz})(1-\delta). \qquad \text{(A4)}$$

Since $1>\delta>0,\theta>0,\theta_z>0,\theta_{zz}<0,f\geq 0,f_z>0,f_{zz}>0,\bar{c}\geq 0,\bar{c}_z>0,\bar{c}_{zz}>0$ by assumption, and $\delta-(f+\bar{c})(1-\delta))>0$ by Proposition 2.A1, then z_i^* is a maximum. Since (A3) is independent of i, it means that $\forall i,j$ $z_i^*=z_j^*$, which implies that all players have a common optimum or z^* is obtained by solving (A2).

Proof of comparative statics on z^*. Rewriting (A3), let $F=\theta_z(\delta-(f+\delta\bar{c})(1-\delta))-\theta(f_z+\delta\bar{c}_z)(1-\delta)$. By the implicit function theorem and (A4), for any parameter w, we have the general result that $sign[\frac{\partial z^*}{\partial w}]=sign[\frac{\partial F}{\partial w}]$. Thus, we have: *(i)* $sign[\frac{\partial z^*}{\partial \bar{c}}]=sign[-\theta_z\delta(1-\delta)]\Rightarrow\frac{\partial z^*}{\partial\theta}<0$; *(ii)* $sign[\frac{\partial z^*}{\partial\theta}]=sign[-(f_z+\delta\bar{c}_z)(1-\delta)]\Rightarrow\frac{\partial z^*}{\partial\bar{c}}<0$; *(iii)* $sign[\frac{\partial z^*}{\partial f}]=sign[-\theta_z(1-\delta)]\Rightarrow\frac{\partial z^*}{\partial f}<0$.

APPENDIX 2: A NOTE ON INCENTIVES FOR COORDINATED PUNISHMENTS

As we note, our focus here is on the 'best' case for punishments to create self-enforcing, cooperative federations. Although we reserve the analysis of coordination problems for later work, to provide some indication of how the states might have incentives to coordiante, we sketch some indicative results here.

Suppose the center induces a state *j* to exit in period *t-1*. Since *S* is the surplus under the fully cooperative equilibrium (or, alternatively, $n(x^{u*}-x^{L*})$), then let S_{-j} indicate the surplus without *j*. Solving for $S_{-j}-S$ we have that $S_{-j}\geq S$ iff

$$\theta\theta_{-j}(1-\delta)(\bar{c}_{-j}-\bar{c})+(\theta_{-j}-\theta)+(\theta\bar{c}_{-j}-\theta_{-j}\bar{c})(1-\delta)\geq 0 \qquad \text{(A1)}$$

where the subscripted terms indicate the values in the reduced federation and the non-subscripted terms the values in the full federation. Using this result, we can turn to an examination of when the reduced federation will be sustainable given the previous equilibrium conditions. To meet this criterion, both the states and the center are made no worse off (and therefore have strong incentives to enforce the previous bargain) under the reduced federation versus the full federation. This is a minimal but illuminating condition of punishment coordination.

(A1) contains two effects on the size of the surplus. On the one hand, the surplus decreases in the smaller federation from decreased scale, in other words since $\theta(n-1,z)<\theta(n,z)$. Second, the surplus increases if exit costs of the eliminated state are *higher* than the average exit costs of the full federation, since exit costs decrease the surplus. If the second effect is dominated by the first effect then the surplus increases (i.e. $S_{-j}\geq S$). If the first dominates the second *or* if the exit costs of *j* are *lower* than the average exit costs in the full federation, then the surplus decreases (i.e. $S_{-j}<S$).

This suggests three interesting cases to examine. Consider first two cases in which $S_{-j}<S$. If , $S_{-j}-\sum_{i\neq j} x_i*<0$, then then there is *no profile* of sustainable, or incentive compatible, payouts such that the both the states can remain rent neutral and the center will not continue to unravel the federation. Here, the size of the existing payouts is sufficiently close to the boundary of the constraint the center puts on the size of the payouts (in other words the upper limit on average payouts x^{u*}), that the decrease in the surplus is greater than the "excess rent" paid to the center. A second possibility is that $S_{-j}-\sum_{i\neq j} x_i*>0$ when . In this case, the center will take the action if and only if its rents from excluding the incremental state are sufficiently low. In other words, if

$$f+c_j+1+\sum_{t=1}^{\infty}\delta^t(S_{-j}-\sum_{i\neq j}x_i^*)\geq\sum_{t=0}^{\infty}\delta^t(S-\sum_i x_i^*)$$

Next, note that the right hand side can be decomposed into its components $\sum_{t=0}^{\infty}\delta^t(S_{-j}-\sum_{i\neq j}x_i^*)+\sum_{t=0}^{\infty}\delta^t(1-x_j^*)$, which yields the result that the center will be better off iff

$$x_j^* \geq \delta-(1-\delta)(f+c_j)\, x_j^* \geq \delta-(1-\delta)(f+c_j). \qquad \text{(A2)}$$

(A2) captures the intuition that if the ongoing rent the center earns is sufficiently large (in other words if its equilibrium payoff to that state is relatively low), it will prefer to keep that state in. If on the other hand, the payout to that state is large relative to what the center can earn by a one-period deviation forcing state *j* to exit, it will have an incentive to force that state out. In this sense, therefore, (A2) states that if a state is getting a large rent relative to its exit costs, then the center will be able to gain while leaving the other states rent neutral. This implies that adding the chance for the center to selectively punish will force a "fairness" on the sustainable divisions in which the stronger (or lowest exit cost) states will get the highest rent relative to the weaker (higher exit cost) states.

If the surplus under the reduced federation is larger than under the full federation the center has a strong incentive to eliminate the state. If the incremental surplus can be captured by the center, each of the remaining states can remain rent-neutral. In this case, the center is strictly better off by inducing one state to leave and moving toward a higher rent position for itself. This points to an approach to identifying "equilibrium federations"–in other words, given the characteristics of the states, how will states sort themselves into appropriate institutional arrangements–which we undertake elsewhere.

Table 6.1: 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.

References

Alesina, Alberto, and Enrico Spolaore. 1997. "On the Number and Size of Nations." *Quarterly Journal of Economics*

Alesina, Alberto, Enrico Spolaore, and Romain Wacziarg. 2000. "Economic Integration and Political Disintegration," *American Economic Review*, 90 (5) 1276-1296.

Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.

Bahl, Roy, and Christin Wallich. 1992. "Intergovernmental Fiscal Relations in China," Working Papers, Country Economics Department, The World Bank, WPS 863.

Bednar, Jennifer L. 1994. "The Federal Problem, MS, Stanford University.

Bednar, Jennifer L. 1996. "Federalism: Unstable by Design," MS, Stanford University.

Bednar, Jennifer L., William Eskridge, and John A. Ferejohn. 2001. "A Political Theory of Federalism," in John Ferejohn, Jack Rakove, Jonathan Riley, eds., *Constitutions and Constitutionalism*. New York: Cambridge University Press.

Bednar, Jennifer L., John A. Ferejohn, and Geoffrey 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.

Blanchard, Olivier, and Andrei Shleifer. 2000. "Federalism with and without Political Centralization: China vs. Russia." W.P. M.I.T.

Calvert Randall L. 1995. "Rational Actors, Equilibrium, and Social Institutions," in Jack Knight and Itai Sened, eds., *Explaining Social Institutions.* Ann Arbor: University of Michigan Press.

Caves, Richard E. 1972. *American Industry: Structure, Conduct, Performance, 3rd Edition*. Englewood Cliffs, NJ: Prentice Hall.

Chen, Yan, and Peter C. Ordeshook. 1994. "Constitutional Secession Clauses." *Constitutional Political Economy* 5:1, pp. 45-61.

Cremer, Jacques, and Thomas Palfrey. 2000. "Federal Mandates by Popular Demand," *Journal of Political Economy* (October) 108:1 905-27.

de Figueiredo, Rui J. P., Jr., and Barry R. Weingast. 2001a. "Pathologies of Federalism, Russian Style: Political Institutions and Economic Transition" Paper prepared for delivery at the Conference, "Fiscal Federalism in the Russian Federation: Problems and Prospects for Reform," Higher School of Economics, Moscow, Russia, January 29-20, 2001.

de Figueiredo, Rui J. P., Jr., and Barry R. Weingast. 2001b. "Constructing Self-Enforcing Federalism in the Early United States and Modern Russia " Paper prepared for delivery at the Conference, "Fiscal Federalism in the Russian Federation: Problems and Prospects for Reform," Higher School of Economics, Moscow, Russia, January 29-20, 2001.

Dewatripont, M. and E. Maskin. 1995. "Credit and Efficiency in Centralized and Decentralized Economies." *Review of Economic Studies* 62: 4 (October), pp. 541-55.

Elazar, Daniel J. 1987. *Exploring Federalism*. Tuscaloosa: University of Alabama Press.

Elkins, Stanley, and Eric McKitrick. 1993. *The Age of Federalism: The Early American Republic, 1788-1800.* New York: Oxford University Press.

Ellis, Richard E. 1987. *The Union at Risk: Jacksonian Democracy, States' Rights and the Nullification Crisis*. New York: Oxford University Press.

Fehrenbacher, Don E. 1978. *The Dred Scott Case: Its Significance in American Law and Politics*. New York: Oxford University Press.

Fehrenbacher, Don E. 1980. *The South and the Three Sectional Crises*. Baton Rouge: Louisiana State University Press.

Freehling, William W. 1966. *Prelude to Secession: The Nullification Controversy in South Carolina, 1816-1836.* New York: Harper.

Freehling, William W. 1990. *The Road to Disunion.* Vol. I: "Secessionists at Bay, 1776-1854." New York: Oxford University Press.

Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory*. Cambridge: M.I.T. Press.

Garman, Christopher, Stephan Haggard, and Eliza Willis. 1999. "Fiscal Decentralization A Political Theory with Latin American Cases" W.P. UCSD.

Gibbons, Robert. 1992. *Game Theory for Applied Economists*. Princeton, NJ: Princeton University Press.

Gibbons, Robert, and Andrew Rutten. 1996. "Hierarchical Dilemmas: Social Contracts with Self-Interested Rulers, " MS, Cornell University.

Green, Edward, and Robert Porter. 1984. "Noncooperative Collusion Under Imperfect Price Information." *Econometrica*, pp. 87-100.

Greif, Avner. 1997. "Self-enforcing Political Systems and Economic Growth: Late Medieval Genoa," in Robert Bates, Avner Greif, Margaret Levi, Jean-Laurent Rosenthal, and Barry R. Weingast, *Analytic Narratives*. Princeton: Princeton University Press.

Greif, Avner. 2000. *Culture and the Institutional Foundations of States and Markets: Historical and Comparative Institutional Analysis of Genoa and the Maghribi Traders*. New York: Cambridge University Press (forthcoming).

Greif, Avner, Paul Milgrom, and Barry R. Weingast. 1994. ""Commitment, Coordination, and Enforcement: The Case of the Merchant Guilds," *Journal of Political Economy*. 102: 745-76.

Hardin, Russell. 1989. "Why a Constitution?" in Bernard Grofman and Donald Wittman, eds., *The Federalist Papers and the New Institutionalism*. New York: Agathon Press.

Holt, Michael F. 1999. *The Rise and Fall of the American Whig Party.* New York: Oxford University Press.

Inman, Robert P., and Michael Fitts. 1990. "Political Institutions and Fiscal Policy: Evidence from the U.S. Historical Record," *Journal of Law, Economics, and Organization*. 6 (special issue): 79-132.

Inman, Robert P., and Daniel L. Rubinfeld. 1997. "The Political Economy of Federalism," in Dennis C. Mueller, ed., *Perspectives on Public Choice Theory*. New York: Cambridge University Press.

Jones, Mark P., Pablo Sanguinetti, and Mariano Tommasi. 2000. "Politics, Institutions, and Fiscal Performance in a Federal System: An Analysis of the Argentine Provinces." *J. of Development Economics*.

Laffont, Jean-Jacques, and Jean Tirole. 1988. "The Dynamics of Incentive Contracts." *Econometrica* 56:5, pp. 1153-75.

Knupfer, Peter B. 1991. *The Union as It Is: Constitutional Unionism and Sectional Compromise, 1787-1861.* Chapel Hill: University of North Carolina Press.

Kreps, David. 1990. *A Course in Microeconomic Theory*. Princeton, NJ: Princeton University Press.

Kreps, David M., Paul Milgrom, John Roberts, and Robert B. Wilson. 1982. "Reputation and Imperfect Information." *Journal of Economic Theory* 27, pp. 253-279.

Lenner, Andrew C. 2001. *The Federal Principle in American Politics, 1790-1833.* Lanham, MD: Rowman and Littlefield.

Lijphart, Arend. 1984. *Democracies: Patterns of Majoritarian and Consensus Government in Twenty-One Countries*. New Haven: Yale University Press.

McKinnon, Ronald I. 1997. "Market-Preserving Fiscal Federalism in the American Monetary Union," in Mario I. Blejer and Teresa Ter-Minassian, eds., *Macroeconomic Dimensions of Public Finance*, New York: Routledge, pp.73-93.

Meinig, D.W. 1993. *The Shaping of America: A Geographic Perspective on 500 Years of History. Vol 2, Continental America, 1800-1867*. New Haven, Yale University Press.

Middlekauff, Robert. 1982. *The Glorious Cause: The American Revolution, 1763-1789.* New York: Oxford University Press.

Milgrom, Paul R., Douglass North, and Barry R. Weingast. 1990. "The Role of Institutions in the Revival of Trade: The Medieval Law Merchant, Private Judges, and the Champagne Fairs," *Economics and Politics* (March) 2: 1-23.

Milgrom, Paul R., and John Roberts. 1990. "Bargaining and Influence Costs." In James Alt and Kenneth A Shepsle, eds., *Perspectives on Positive Political Economy*. New York: Cambridge University Press.

Montinola, Gabriella, Yingyi Qian, and Barry R. Weingast. 1995. "Federalism, Chinese Style: The Political Basis for Economic Success in China" *World Politics* (October) 48: 50-81.

Moore, Glover. 1953. The Missouri Controversy: 1819-1821. [no city]: University of Kentucky Press.

Morgan, Edmund S. 1977. *The Birth of the Republic: 1763-89*. Chicago: University of Chicago Press. Revised ed.

Oates, Wallace. 1992. *Fiscal Federalism*. New York: Harcourt Brace Jovanovich.

Oi, Jean. 1992. "Fiscal Reform and the Economic Foundations of Local State Corporatism in China," *World Politics* (October) 45: 99-126.

Oksenberg, Michel, and James Tong. 1991. "The Evolution of Central-Provincial Fiscal Relations in China, 1971-1984: The Formal System." *China Quarterly*, March.

Ordeshook, Peter C., and Olga Shvetsova. 1997. "Federalism and Constitutional Design," *Journal of Democracy* (1997) 8(1): 27-42.

Osborne, Martin J., and Ariel Rubenstien. 1994. *A Course in Game Theory*. Cambridge, MA: MIT Press.

Persson, Torsten, and Guido Tabellini. 1996. "Federal Fiscal Constitutions: Risk Sharing and Moral Hazard," *Econometrica* 64:3, pp. 623-646.

Persson, Torsten, and Guido Tabellini. 1996. "Federal Fiscal Constitutions: Risk Sharing and Redistribution." *Journal of Political Economy* 104:5, pp. 979-1009.

Poterba, James, and Jürgen von Hagen, editors. 1999. *Fiscal Institutions and Fiscal Performance*. University of Chicago Press.

Rakove, Jack N. 1996. *Original Meanings: Politics and Ideas in the Making of the Constitution*. New York: Knopf.

Rakove, Jack, Andrew Rutten, and Barry R. Weingast. 2000. “Ideas, Interests, and Credible Commitments in the American Revolution” Working Paper, Hoover Institution.
Riker, William H. 1964. *Federalism: Origins, Operations, and Significance*. Boston: Little, Brown.
Riker, William H. 1987. “The Lessons of 1787.” *Public Choice* (September) 55, pp. 5-34..
Rodden, Jonathan. 1999. “Strategy and Structure in Decentralized Fiscal Systems: A Comparative Theory of Hard Budget Constraint.,” Working Paper, Yale University.
Rodden, Jonathan. 2000. “The Dilemma of Fiscal Federalism: Hard and Soft Budget Constraints Around the World,” Working Paper, MIT.
Rubinfeld, Daniel. 1987. "Economics of the Local Public Sector,” in A. J. Auerbach and M. Feldstein, eds., *Handbook of Public Economics, vol. II.* New York: North-Holland.
Shirk, Susan. 1993. *The Political Logic of Economic Reform in China*. Berkeley: University of California Press.
Schlesinger, Arthur Meier. 1922. "The State Rights Fetish," in his, *New Viewpoints in American History*. New York: Macmillan.
Solinger, Dorothy. "The Floating Population as a Form of Civil Society , II Paper for 43d Annual Meeting of the Association for Asian Studies, New Orleans, April 11-14, 1991.
Solnick, Stephen L. 1998. *Stealing the State: Control and Collapse in Soviet Institutions*. Cambridge: Harvard University Press.
Stepan, Al. 1998. [MS on top down federalism] **
Sydnor, Charles S. 1948. *The Development of Southern Sectionalism, 1819-1848.* Baton Rouge, Louisiana State University Press.
Tiebout, Charles. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy*, 64:416-24.
Tirole, Jean. 1988. *The Theory of Industrial Organization*. Cambridge: M.I.T. Press.
Tomassi, Mariano . 2000. A Principal-Agent Building Block for the Study of Decentralization and Integration. Working Paper, Universidad de San Andres, Buenos Aires.
Treisman, Daniel. 1999. “Political Decentralization and Economic Reform: A Game Theoretic Analysis,” *American Journal of Political Science*. (April) 43:2, pp. 488-517.
Treisman, Daniel. 1999. “Corruption and Federalism,” Working Paper. UCLA.
Treisman, Daniel. 2000. *After the Deluge: Regional Crises and Political Consolidaton in Russia*. Ann Arbor: University of Michigan Press.
Weingast, Barry R. 1995. "The Economic Role of Political Institutions: Market-Preserving Federalism and Economic Development," *Journal of Law, Economics, and Organization*. (Spring) 11: 1-31.
Weingast, Barry R. 1997. "The Political Foundations of Democracy and the Rule of Law," *American Political Science Review*. (June) 91: 245-63.
Weingast, Barry R. 1998. “Political Institutions and Civil War: Institutions, Commitment, and American Democracy,” in Robert Bates, Avner Greif, Margaret Levi, Jean-Laurent Rosenthal, and Barry R. Weingast, *Analytic Narratives*. Princeton: Princeton University Press.
Williamson, Oliver. ed. 1990. *Industrial Organization*. Brookfield, VT: Edward Elgar.
Wong, Christine P.W. 1991. "Central-Local Relations in an Era of Fiscal Decline: The Paradox of Fiscal Decentralization in Post-Mao China." *China Quarterly*, December: pp. 691-715.

Wood, Gordon. 1969. *The Creation of the American Republic, 1776-1787.* New York: Norton.
Wood, Gordon. 1991. *Radicalism and the American Revolution*. New York: Vintage Books.