Self-Enforcing Federalism

Rui J. P. de Figueiredo, Jr.
University of California, Berkeley

Barry R. Weingast
Stanford University

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 subunits, 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 trade-offs among the benefits from central goods provision, the ability of the center to impose penalties for noncompliance, 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 (see, e.g., Fearon, 2000; Calvert, 1995; Greif, 1997, 2001; Hardin, 1989; Milgrom, North, and Weingast, 1990: Ordeshook, 1992: Przeworski, 1991, 2000; and Weingast, 1997). 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.

The authors gratefully acknowledge the helpful comments of Jenna Bednar, Jonathan Bendor, Mel Bernstein, Scott Gehlbach, Natalia Ferretti, Ed Green, Douglas Grob, Dan Kelemen, Robert Powell, Barak Richman, Thomas Romer, Pablo Spiller, Ken Shepsle, and Oliver Williamson and seminar participants at Harvard University, Princeton University, Stanford University, University of California–Berkeley, University of California–Los Angeles, University of California–San Diego, San Diego University, and Yale University. In addition, the authors are also grateful for excellent comments from two anonymous reviewers. All errors are solely the fault of the authors.

The Journal of Law, Economics, & Organization, Vol. 21, No. 1,
doi:10.1093/jleo/ewi005

In this article 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.

1.1 The Twin Dilemmas of Federalism

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 overawing 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 (Riker, 1964). Further, since constitutions are typically 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" (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). 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 trade-off*: 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 that is too strong, a federation typically fails because the national government compromises state independence. Reflecting this trade-off, several theorists emphasize federalism's instability (Riker, 1964; Bednar, 1996; Ordeshook and Shvetsova, 1997; Bednar, Eskridge, and Ferejohn, 2001).

1.2 Motivating Examples: The Dilemmas in Action

Throughout history, many federations have experienced the fundamental trade-off during federal crises. The cases provide important illustrations of these trade-offs and how they might be resolved, to the extent they have been.

1.2.1 Creating American Federalism.

In the American case, numerous debates have centered on the benefits and dangers of centralization. Nowhere was the tension more vividly demonstrated than in the debates following the American Revolution. The principal criticism of the Articles of Confederation by Federalist leaders was that the national government had insufficient institutional power to supply critical national public 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.[1] The Federalists believed that the national government should be granted strong taxation powers in order to have resources to achieve these ends (Morgan, 1977:chap. 9; Kaplanoff, 1991). Some Anti-Federalists admitted a concern about the undersupply 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.[2]

Under the Articles of Confederation, the Anti-Federalists' political power allowed them to maintain the balance in their favor. The failure to provide the national government with sufficient authority to finance common defense, police internal trade, and control currency led to a myriad of common pool problems: states refused to contribute to national finances, many erected trade barriers, and currencies were "oversupplied" by some states (Middlekauff, 1982; Kaplanoff, 1991).[3]

Of interest is that the failure to provide adequate authority, which led to the exacerbation of the first dilemma, was caused by concerns about the second: one of the main problems with the Federalist proposals to address the problems under 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 or exploited. This background sets the stage for the Federalist resolution of these problems with the new Constitution in 1787.

1.2.2 Latin American Federations. As in the American case, many Latin American federations have exhibited sharp trade-offs in designing effective institutions. In contrast to the American cases, however, many Latin American federations have resulted in degenerate federal arrangements. In Mexico, for example, the central government historically provided states with 80% (or more) of their revenue. Along with the revenue comes the national government's rules and restrictions. Further, in many Latin American states, the national government, not the local government, remains the locus of regulatory control over the economy. Any attempt to ignore the rules risks the withdrawal of all funds. For example, consider the rise of political competition to the PRI, the political party that has dominated Mexico since their revolution. The national government frequently punishes local governments captured by the

1. 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).

2. 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."

3. 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).

political opposition by withdrawing budgetary funds (Diaz-Cayeros, Magaloni, and Weingast, 2004). This sets a very high price on voting for the opposition for local citizens. In combination, these features of federalism imply that states are not autonomous, sovereign entities, as federalism requires. Instead, they remain administrative agents of the national government.

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

This legacy set the stage for the political conflict in modern Russia. Of interest is that many observers have commented on various parts of the breakdown in Russian federalism (e.g., Ordeshook and Shvetsova, 1997; Solnick, 1998; Blanchard and Shleifer, 2000; de Figueiredo and Weingast, 2000): some focus on the lack of constraints on the center, others that a too weak center cannot control runaway and profligate states. Through the lens of the fundamental trade-off, however, the problems are related. Having succeeded the Soviet Union for sovereignty over its territory, Russia under Yeltsin has at once the problem of too weak a center and too strong a center. On the one hand, the Russian government today has only a limited ability to commit to limits on its behavior. In particular, states lack a consensus about the appropriate limits on the national government.

Perhaps ironically, but certainly consistent with the trade-off in Russia, unlimited governmental authority ultimately has led to the weakening of the federal apparatus, including the central government, as the failure to define limits on the center led to less willing delegation of authority to and compliance with the center by the states. This has led to a central government that is financially weak. Financial weakness, in turn, has allowed many local and regional governments to grab considerable de facto independence. Short of announcing sovereignty, as in Chechnya, the central government has acquiesced to most of these assertions of regional government power. Together, therefore, the problem is not simply one of too weak or too strong a center, but appropriately aligning incentives for the proper uses of authority and compliance with agreements.

In many cases, then, the twin dilemmas have created crises in federal institutional arrangements. In this article, we attempt to explore both the implications of these trade-offs and how they might be resolved, if at all.

1.3 The Literature on Federalism

Three rich streams of the literature relate to the two fundamental dilemmas of federalism.[4] The first and largest stream studies the problem of state shirking

4. These three literatures focus on aspects of endogenous federalism. In addition, there is a much larger literature on the effects of federalism, dominated by economists (such as Tiebout, 1956; Oates, 1972, Rubinfeld, 1987). There is also a political science literature on the effects of federalism on various problems, such as ethnic conflict (see Lijphart, 1984:chap.10), budget deficits (Rodden, 1999, 2000; Poterba and von Hagen, 2000), and corruption (Treisman, 1999).

and common pool problems from subnational governments. The settings vary dramatically, including demand for federal spending; budgets, state borrowing, soft budget constraints, and deficits; and voting (see, e.g., Bednar, 1998a,b; Bednar, Eskridge, 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, 1996a,b; Poterba and von Hagen, 2000; Rodden, 1999, 2000; and Sanguinetti, 1995). 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. Riker (1964) and Bednar (1996), for example, examine how central governments tend to expand their powers over time. Chen and Ordeshook (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 article 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 articles 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.

1.4 Overview of the Argument and Plan of the Article

The model we develop here has a number of purposes. By clearly defining the components of the institutional design problem, we more precisely elaborate when a set of federal institutions can be sustained. In our formulation, we use the game theoretic concept of equilibrium as a way of studying the sustainability of a federation. In this way, the model helps to clarify a number of critical features of the design of federal systems which will determine their stability. Using this analysis, we then analyze how these equilibria change with respect to the exogenous characteristics of the players attempting to capture the benefits of federation and the opportunities they are attempting to capture. Finally, as we show, in many situations, many possible institutions can sustain a federation. Our model helps to clarify, in these situations, which specific choices will be made by the players who design a federal constitution. By doing so, we are able

to examine the way federal institutions become stronger or weaker, depending on a number of important features of the interaction between various interested parties. In this sense, we are able to generate a number of predictions about the characteristics of federations designed in different contexts.

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 among subnational 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 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; Milgrom, North, and Weingast, 1990; Persson and Tabellini, 1996a; Bednar, 1996). If the moral hazard problem by the subunits 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 trade-off: 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. The model endogenizes several aspects of federal authority: the degree of state participation and shirking, and limits on the federal government. Using this framework, 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 a wide range of federations, including the 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: the weaker the set of states, the more productive the center, and the larger the federation, ceteris paribus, the weaker is central authority in equilibrium.

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; Bolton and Roland, 1997; Alesina, Spolaore, and Wacziarg, 2000), our model shows that there is a trade-off between heterogeneity and scale. The difference, however, is that we show that this trade-off 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.

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 trade-offs; there

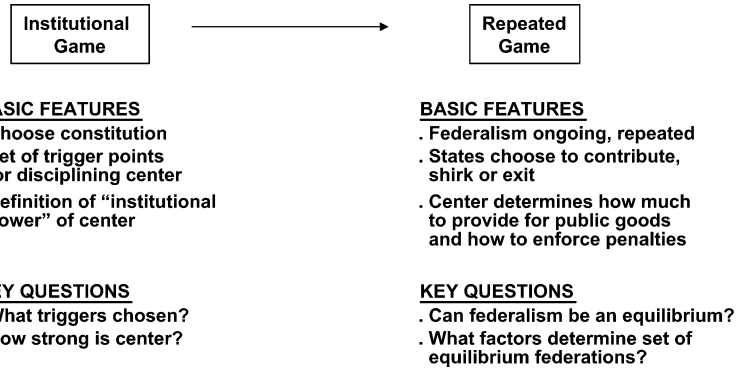

Figure 1. Modeling Overview.

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 1, is first to solve the characteristics of the federal equilibrium *given* the institutions of the federation, and then to study 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 $i = 1,\ldots,N$, and a *central government* called C.[5] The sequence of moves is shown in Figure 2. First, the states choose one of three actions $A = \{P, S, E\}$ for contribute, shirk, and exit. If a state chooses P, it *contributes* one unit to the center. The indicator variable k_i equals one if a state contributes and zero 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.

5. The center is a distinctive actor in our model. Paralleling most federations, this implies that the set of states do not directly set a nationally governing policy. Although in the central government, officials might also be members of a state, this notion of indirectness is crucial to the basic premise of the two dilemmas: the central government, even when constituted by the states, becomes a self-interested actor. Indeed, as both the Anti-Federalists and the Federalists warned during the American Constitutional debates, the danger is that the center becomes captured by a faction, a subset of the states, or some minority that pursues its own self-interest.

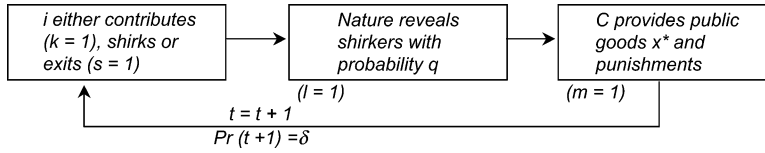

Figure 2. Sequence of Play in RG Stage Game.

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 one 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 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. We intentionally leave the substantive interpretation of the parameter z very broad: the delegation of power to the center vis-à-vis the states. That said, a wide range of constitutional variables affect this power in practice, including for example, relative to the subnational governments, the degree to which the center holds: taxing power; policy control over the economy; control over subnational government budgets; appointment and/or removal power of subnational government officials; and the power to take over a state.

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 disentangle 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 $c_i(z) \geq 0, c_i'(z) > 0, c_i'' \geq (z) \leq 0$. 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, $c_j(z) > c_j(z)$, then $c_i(z) > c_i(z) \forall z$. 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 may fail is that observability of state shirking is imperfect. When states erect nontariff trade barriers or enforce regulations arbitrarily, there are often disagreements on whether violations of federal agreements have occurred. To capture the notion that monitoring of violations is imperfect, the second step in the stage game is that a nonstrategic 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 one for the indicator variable l_i. All players observe only the vector $l = (l_1, l_2, \ldots, l_N)$, 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 national public good in which all states receive the same payoff, or one 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* $\mathrm{x} = (x_1, x_2, \ldots, x_N)$, which is the amount of payments made to each subunit. 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 subunits are modified by $\theta(n,z)$. We assume that a stronger center can better provide certain goods,[6] 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 $\theta(n,z)_n > 0, \theta(n,z)_{nn} \leq 0$. These characteristics of the function $\theta(n,z)$ follow from the notion that there are certain central goods that are more efficiently provided at the central level. These efficiencies can arise either because of greater economies of scale or greater coordination. Take, for example, two of the primary public goods that national governments provide within a federation: defense and a common market. In both of these, the substantial fixed costs of defense infrastructure and regulatory infrastructure suggest increasing returns to scale.[7]

C also chooses a *punishment or extraction strategy* $\boldsymbol{m} = (m_1, m_2, \ldots, m_N)$, which is a vector of indicators designating if an additional fee $f(z)$ will be levied against each subunit 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

6. We use the term "central goods" to define the product of the center, since our model allows for both public and nonpublic 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 nondiscriminatory 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).

7. Although such economies might be better thought of in terms of the size of populations and not the number of states, if we fix the population of any state, then for those areas where such economies exist, there will be an increase in the efficiency of the central goods provision as the number of states increases. With respect to the parameter z, in order to provide these central goods, the center needs power and authority. The specification that $\theta(n,z)$ is increasing in the grant of these powers is a natural extension of the discussion in Section 1.2. For example, providing the national government with the power to determine taxes, but not providing it with the resources and power to enforce tax policy, will make the provision of central goods much less effective. Indeed, this was one of the central features of the American experience under the Articles of Confederation.

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

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

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 the 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, $c_i(z)$.[8,9]

8. 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. While this is primarily for convenience, 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 be 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.

9. 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 its 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.

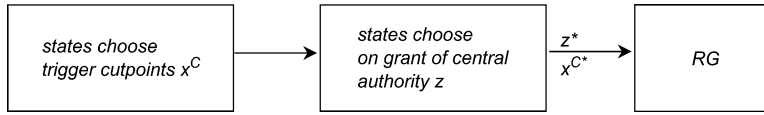

Figure 3. Sequence of Play in *IG*.

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

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

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 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 that gives the subnational 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 subnational 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. Returning to our original question, if cooperative outcomes can be sustained in equilibrium, then the federal arrangements are deemed to be sustainable.

Cooperative equilibria are defined as those in which, on the equilibrium path, all states choose P 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 the *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* $\mathbf{x}^*$, *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 cease cooperation and exit; if any state exits, the center will set* $\mathbf{x} = 0$ *and* $\mathbf{m} = \mathbf{1}$.

The grim trigger strategy says that, if ever a player deviates from the cooperative equilibrium, all players irrevocably enter a defection stage. Under the grim trigger, the players will cooperate only as long as all the other players have always cooperated.[10] Notably, for convenience, the defection strategy that we examine is a simple one—all states cease cooperation and possibly exit.

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, 1994:274–75). 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 in Bendor and Mookherjee

10. Notably, for convenience, the defection strategy that we examine is a simple one—all states have two options: cease cooperation and stay in the federation or completely exit. Although this simplification might seem extreme, the key feature we wish to highlight is that the *threat* of either secession or noncooperation has to be credible. Such threats of either noncooperation or secession have indeed been observed in a number of cases. Quebec, Catalonia, and certain Russian regional governments have all used threats of secession to extract gains from federations. Similarly, in the antebellum United States, groups of states simply ceased cooperating with the central government (short of secession), voting negatively on most major initiatives and hence holding up the national government. Indeed, South Carolina's attempt to *nullify* federal policies was a prime example of such noncooperation.

(1987); see also Gibbons (1992)]. While analytically more convenient, GT yields substantively similar results to any other strategy in this class.[11]

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

Proposition 1. Fix δ,z,n. *If*
(Assumption 1) shirking constraint: $f(z)q(z) > 1$, and
(Assumption 2) gains from federation: $\theta(n,z) > 1$,

then there exist GT equilibria in which

(*i*) *states' cooperation threshold:* $x_i^* \geq (1-c_i(z)(1-\delta))/(\theta(n,z)) = x_i^{L^*} \forall i$;
(*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. See the appendix.

Proposition 1 provides a number of insights into the ongoing dynamic between the center and the states and most importantly 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 federation to be an equilibrium: the center must be given a strong enough incentive to detect and punish potential shirkers.[12] Notice that this was precisely the point that the Federalists made in the debates with the Anti-Federalists in the years preceding the adoption of the U.S. Constitution: if the center was not sufficiently strong, the states would simply renege or shirk on the federal bargain. In addition, the condition implies that the constraint is more stringent as the function $q(z)$ becomes smaller. The reason is that, the lower $q(z)$ is, the higher $f(z)$ must be in order for $f(z)q(z)$ to be greater than one. 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. Indeed, returning to the example from the early history of the United States, this was the counterpoint made by the Anti-Federalists in the face of requests for a stronger central government: only in situations in which the center did not need to be

11. 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 cooperative punishment in the appendix.

12. 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.

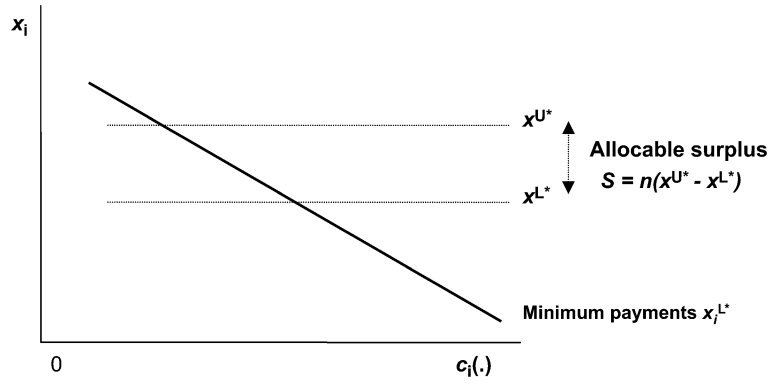

Figure 4. Equilibria of the *RG*.

"too strong" to enforce the federal agreement should the center have such power. Together, these two results are a theoretical encapsulation of the fundamental trade-off in designing federal institutions, and indeed, the twin results show why, in many cases, when the conditions cannot be jointly satisfied, a federal system will not be sustainable.

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 Assumptions 1 and 2 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 this threshold x_i^L* is decreasing in $c_i(z)$, the minimum amount required to provide an incentive for a state to remain in the federation falls as the exit costs rise. Similarly, condition (ii) defines the center's cooperation threshold x^U*. It states that the center must not be asked to return so much (more than this threshold) to the states on average, 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 "take the money and run," that is, taking all the contributions in the current period for itself even though this implies losing all future payments. Maintaining the federation requires that the 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 4. The states are arrayed continuously along the horizontal axis by their level of exit costs. The heavy line shows the minimum level of contributions

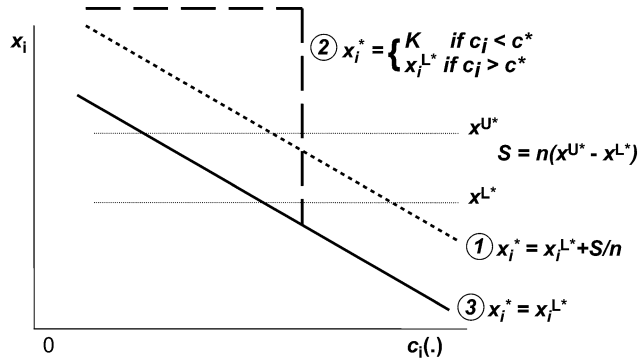

Figure 5. Illustrative Surplus Allocations.

a state is willing to accept before exiting x_i^L*. The line slopes downward since this quantity decreases as exit costs increase. The lower dashed line 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** that 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 5 illustrates this point.[13] It shows three possible equilibrium profiles, x, all of which are consistent with the conditions in

13. 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 Mookherjee, 1987; Weingast, 1997) which we reserve for later work. In Appendix B, we provide an analysis of these issues.

Proposition 1: (1) the allocation of the surplus equally among the states (i.e., 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 subunits. This indicates that exit costs potentially shift rents 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.[14] In terms of our more general question, this analysis indicates how flexibly and loosely, and how easily a federal arrangement that is sustainable can be achieved. With respect to the discount factor, as the players value the future more, more profiles can be sustained in equilibrium (i.e., $(\partial S/\partial \delta) > 0$; all proofs of these results are shown in Appendix A). 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., $(\partial S/\partial \theta) > 0$). 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., $(\partial S/\partial f < 0)$. 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.[15] 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 increasing exit costs shift 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. Indeed, Treisman (1999b) and others have observed that the center grants far better deals to those subnational governments that have a credible exit threat.

14. Here we mean how large is the surplus or excess rent and therefore the set of possible equilibria.

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

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 the center collects but does not return to the states represents an opportunity cost in public benefits foregone.

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, the states' inability to *coordinate* on a punishment strategy means 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 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 $\mathbf{x}$ 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 preplay bargaining power in order to determine the division of rents. Thus we designate the vector $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)$ 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.[16] 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

16. This concept means that, conditional on the existence of an equilibrium to the RG, we characterize the core of ($\mathbf{x}$, z).

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.[17] 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 the following two conditions hold:

(Assumption 3) positive surplus: $\delta - (f(z)+\bar{c}(z))(1-\delta) \geq \frac{1}{\theta(n,z)}$;
(Assumption 4) no shirking: $\mathrm{f}(z)q(z) > 1$.

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

(i) optimal trigger: $x_i^* = (1-\alpha_i, n)\frac{1}{\theta(z^*)}+\alpha_i n(\delta - (f(z^*)+\bar{c}(z^*))(1-\delta))$;
(ii) optimal central power: z^* *solves* $\frac{\theta_z}{\theta} = ((f_z+\delta\bar{c}_z)(1-\delta))/(\delta - (f+\delta\bar{c})(1-\delta))$ and has a unique solution;
(iii) all states participate.[18]

Proof. See the appendix.

Before turning to the implications of the analysis, consider Assumptions 3 and 4. Assumption 3 simply states that S is positive, so an equilibrium can exist. Assumption 4 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 the 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

17. It is useful to clarify that whereas in the previous section both $c(.)$ and z where exogenous, in this section z is endogenous but the function c remains exogenous. Thus whereas the states can choose exit costs given c they cannot influence c itself. In the later discussion, we consider the effect of different levels of c, in other words, what happens as the exogenous function which maps z into exit costs shifts?

18. Note that 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.

a coordination problem: if the definition of central transgression is unarticulated—for example, if the states disagree about the appropriate definition of a transgression—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 Hardin, 1989; Chen and Ordeshook, 1994; and Weingast, 1997).[19]

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 the states 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 in bottom-up federations. 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 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 $(1/\theta(z^*))$. 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 is $\theta(z^*)x_i^* - 1 \geq 0$. Figure 6 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 $x_i^L* = 1 - \bar{c}(z)(1 - \delta))/(\theta(z)$, represented in the figure by the heavy solid line. With the participation constraint, however, every state must receive at least $1/\theta(z^*)$ 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.

19. There are a number of examples of such "brightline" rules about what the national government can and cannot do. In the United States, for example, the 10th Amendment, the prohibition of any official religion and the clear delineation of certain policies in the hands of the states—definition of property rights, regulation of intrastate, commerce and the enforcement of contracts—are all examples of such clearly defined boundaries between center and states. In practice, the main point about these rules is that they make observations of violations easy for enforcement authorities such as the courts and the states themselves.

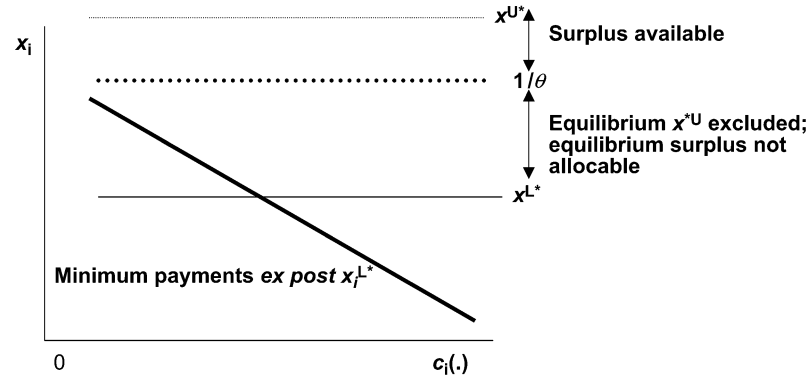

Figure 6. Equilibrium Profiles for the *IG*.

Unlike in the RG, in which participation is fixed, each state's minimum return here is identical.

This raises the problem of the center's commitment to the federal bargain: if once in the federation the center can extract even more, how could it credibly commit not to extract the maximum amount such that a state would not want to leave? And if this commitment problem exists, why would a state "believe" ex ante the promise of the center to provide at least $(1/\theta(z^*))$? The model provides a twofold answer. First, from the point of view of the model, there are multiple equilibria that can be sustained. In some, the center will obtain more, in others the states will obtain more. The question therefore is why are equilibria in which the center is not extracting its maximal amount an equilibrium. The answer comes in two parts. On the one hand, the states themselves can credibly threaten to enforce the equilibrium as the off-path noncooperative punishment is itself a Nash equilibrium.[20] On the other hand, the center has an incentive to maintain the higher level of public goods provision because the triggers exercised here are joint triggers, the proposition shows that the expected punishments from all of the states for a central transgression are sufficient to prevent the center from acting capriciously. Second, from a substantive point of view, this is precisely one of the features of specific institutions which are so crucial—if the institutions do not provide a basis (in part because of the punishments) for the center's commitment, then the suspicious subnational units will not enter the federation. Indeed, this was precisely the basis for caution from the Anti-Federalists in the design of the U.S. Constitution.

Further, returning to our original question of a federation's sustainability, the result means that a federation is even harder to sustain than implied by the results of the previous section.[21] Before, as long as the minimum required to

20. While more complicated notions of group punishments are worthwhile to examine, we leave such an examination to later work. We do provide some initial discussion of these issues in Appendix B.

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

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*.[22]

Fourth, result (ii) in Proposition 2 highlights the central trade-off 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 value of the centrally provided 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 prevention of shirking and central goods provision equal the marginal costs of increased rent extraction.

Notice that in a bottom-up federation, 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 A). This leads to a significant prediction in 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 being 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. The intuition behind both these results is that because the states are concerned about ex post opportunism by the center, they choose weaker central arrangements ex ante, which in turn reduces the ability of the center to provide welfare-enhancing central goods. 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 being equal.

5. Equilibrium Federations

In the previous section, the structure of the federation was fixed and participation was considered only from the perspective of an individual state. 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

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

must an individual state opt into federalism, but 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. 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 nth 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 $1,\ldots,n-1$ 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} similarly for the n federation. Then the n federation dominates the $n-1$ federation iff:

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

Proof. The result follows directly from Proposition 2.

Equation (1) 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, of which our model highlights two types. The first is the direct benefit that the surplus itself, all things being 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.

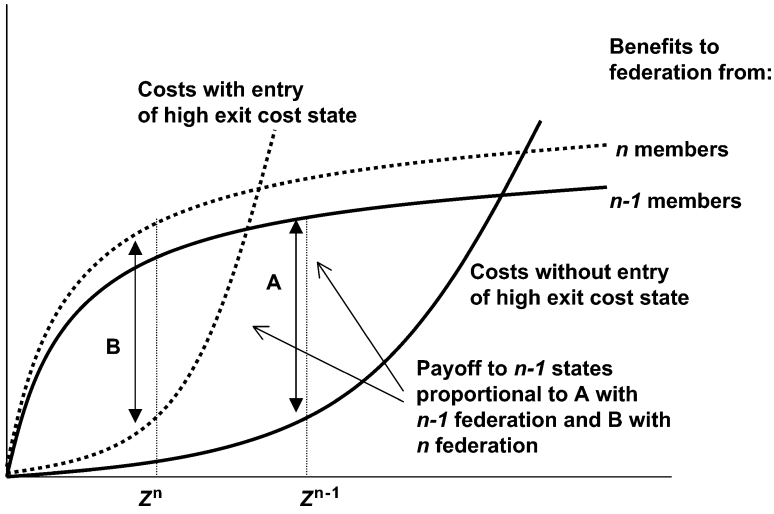

Figure 7. Illustrative Optimal Inclusion of State *n*.

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 Equation (1) 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 the center's incentive to deviate.

We illustrate this trade-off in Figure 7. 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^{n-1}. 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^n. 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., Tiebout, 1956; Alesina and Spolaore, 1997; Bolton and Roland, 1997; Alesina, Spolaore, Wacziarg, 2000): 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—in other words, as the function c shifts upwards for any z—federal institutions will be weaker. In cases such as the European Union, for example, this means that one of two things were happening to prompt expansion to countries with presumably higher exit costs: either the ability of the center to extract rents (e.g., $f(z)$ and $c(z)$) was declining, or the authority of the center was reduced (z) as the federation grew.

6. Conclusion

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 trade-off in the structure of a federation. Institutions designed to address one of the dilemmas tend to exacerbate the other. To be stable, federalism requires a delicate balance of central government powers combined with mechanisms for limiting the center's opportunism.

This article develops a model of self-enforcing federalism, showing how stable federations solve the two fundamental dilemmas of federalism. Our model yields a series of results that highlight which conditions are crucial for the institutions of federalism to be an equilibrium. 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 shift 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 so that 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 to the center is destabilizing.

An important feature of our approach is that the 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. Indeed, as the example of the creation of the U.S. Constitution illustrates, the construction of a set of consensus agreements about the limits on the national government and on state shirking was critical to arriving at a sustainable agreement. 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 we do not examine a model of top-down federations here, our model might be extended to study this form of federalism. A federation designed by the center is likely to leave the center with a greater share of the rents than a bottom-up federation, as we discuss in our model. 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 decrease because of a change in the function c (and not z), 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, Spolaore, and Wacziarg (2000) 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 breakup of nations. We disagree, observing that Alesina, Spolaore, and Wacziarg ignore the endogenous reaction of the center. In response to decreasing 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.

Our article contributes to the growing literature on "equilibrium institutions" (Calvert, 1995; 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 A: Proofs of Propositions Stated in the 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 shirking iff $f(z)q(z) > 1$. Now consider when it will contribute versus exit. If it exits its payoff will be $-c_i(z)$. If it contributes, its expected payoff will be $\sum_{i=0}^{\infty} \delta^t(\theta(n,z)x_i - 1) = \frac{(\theta(n,z)x_i - 1)}{1-\delta}$ 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 to play is $x_i = 0 \forall i$ and $m_i = 1 \forall i$. Thus the payoff to deviating for the center is $\sum_{i=1}^{n} f(z) + \delta c_i(z) + 1 = n(1 + f(z) + \bar{c}(z))$. Its expected payoff to not deviating is $\sum_{t=0}^{\infty} \delta^t \sum_i (1 - x_i) = \frac{n - \sum_i x_i}{1-\delta}$. This in turn implies that the center will stay on the equilibrium path if $\frac{1}{n}\sum_i x_i \leq \delta - (1-\delta)(f(z) + \bar{c})$. 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 a reversion to the Nash equilibrium, enforcement is subgame perfect.

Proof of Comparative Statics in the RG. Note first that $x^{U*} = \delta - (f(z) + \delta\bar{c}(z))(1-\delta), x^{L*} = \frac{1-\bar{c}(z)(1-\delta)}{\theta}$, 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*) $\frac{\partial S}{\partial f} = -(1-\delta) < 0$.

(*iii*) $\frac{\partial S}{\partial \theta} = \frac{1}{\theta}[\delta - (f + \delta\bar{c})(1-\delta)] - \frac{1}{\theta^2}[\delta - (f + \delta\bar{c})(1-\delta) - \theta(1 - \bar{c}(1-\delta)]$. Substituting the expressions for x^U* and x^L*, this simplifies to $\frac{\partial S}{\partial \theta} = \frac{(0-1)x^{U^*} + \theta^{L^*}}{\theta^2} > 0$, since $\theta > 1$

(*iv*) $\frac{\partial S}{\partial \bar{c}} = \frac{(1-\delta)(1-\theta\delta)}{\theta} \rightarrow \frac{\partial S}{\partial \bar{c}} > 0 \delta > \frac{1}{\theta}$.

Proof of Proposition 2. Note first that Assumptions 3 and 4 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 $\max \frac{1}{\theta(n,z,)} + \alpha_i S$ 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

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

we have the condition

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

which implies that for player i, z_i^* solves

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

The second-order condition of Equation (A1) is

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

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 Assumption 3, then z_i^* is a maximum. Since Equation (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 Equation (A2).

Proof of Comparative Statics on z^*. Rewriting Equation (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 Equation (A4), for any parameter w, we have the general result that $\text{sign}\left[\frac{\partial z^*}{\partial w}\right] = \text{sign}\left[\frac{\partial F}{\partial w}\right]$. Thus, we have (*i*) $\text{sign}\left[\frac{\partial z^*}{\partial \bar{c}}\right] = \text{sign}[-\theta_z \partial (1-\delta)] \Rightarrow \frac{\partial z^*}{\partial \theta} < 0$; (*ii*) $\text{sign}\left[\frac{\partial z^*}{\partial \theta}\right] = \text{sign}[-(f_z + \delta \bar{c}_z)(1-\delta)] \Rightarrow \frac{\partial z^*}{\partial \bar{c}} < 0$; (*iii*) $\text{sign}\left[\frac{\partial z^*}{\partial f}\right] = \text{sign}[-\theta_z(1-\delta)] \Rightarrow \frac{\partial z^*}{\partial f} < 0$.

Appendix B: 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 coordinate, 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} indicates the surplus without j. Solving for $S_{-j} - S$, we have that $S_{-j} \geq S$ iff

$$\theta\theta_{-j}(1-\delta)(\bar{c}_{-j} - \bar{c}) + (\theta_{-j} - \theta) + (\theta\bar{c}_{-j} - \theta_{-j}\bar{c})(1-\delta) \geq 0, \tag{A5}$$

where the subscripted terms indicate the values in the reduced federation and the nonsubscripted terms are 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.

Equation (A5) 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 there is no profile of sustainable, or incentive compatible, payouts such that 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 $S_{-j} - \sum_{i \neq j} x_i* > 0$. 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

$$x_j^* \geq \delta - (1 - \delta)(f + c_j) x_j^* \geq \delta - (1 - \delta)(f + c_j). \tag{A6}$$

Equation (A6) 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, Equation (A6) 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 in the penultimate section.

References

Alesina, Alberto, and Enrico Spolaore. 1997. "On the Number and Size of Nations," 112 *Quarterly Journal of Economics* 1027–56.

Alesina, Alberto, Enrico Spolaore, and Romain Wacziarg. 2000. "Economic Integration and Political Disintegration," 90, *American Economic Review* 1276–96.

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

Bahl, Roy, and Christin Wallich. 1992. "Intergovernmental Fiscal Relations in China," working paper, Country Economics Department, The World Bank.

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

Bednar, Jennifer L. 1996. "Federalism: Unstable by Design," Masters thesis, 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," 16 *International Review of Law and Economics* 279–94.

Bendor, Jonathan, and Dilip Mookherjee. 1987. "Institutional Structure and the Logic of Ongoing Collective Action," 81 *American Political Science Review* 131–54.

Blanchard, Olivier, and Andrei Shleifer. 2000. "Federalism with and without Political Centralization: China vs. Russia." working paper, MIT.

Bolton, Patrick, and Gerard Roland. 1997. "The Breakup of Nations: A Political Economy Analysis." 112 *Quarterly Journal of Economics* 1057–1190.

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, 3*rd ed. Englewood Cliffs, NJ: Prentice Hall.

Chen, Yan, and Peter C. Ordeshook. 1994. "Constitutional Secession Clauses," 5 *Constitutional Political Economy* 45–61.

Cremer, Jacques, and Thomas Palfrey. 2000. "Federal Mandates by Popular Demand," 108 *Journal of Political Economy* 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," 62 *Review of Economic Studies* 541–55.

Diaz-Cayeros, Alberto, Beatriz Magaloni, and Barry R. Weingast. 2004. "Tragic Brilliance: Equilibrium Hegemony And Democratization in Mexico." Working Paper, Hoover Institution, Stanford University.

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, MA: MIT Press.

Garman, Christopher, Stephan Haggard, and Eliza Willis. 1999. "Fiscal Decentralization: A Political Theory with Latin American Cases," working paper, University of California San Diego.

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, " Master thesis, Cornell University.

Green, Edward, and Robert Porter. 1984. "Noncooperative Collusion Under Imperfect Price Information," *Econometrica* 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, eds. *Analytic Narratives*. Princeton, NJ: Princeton University Press.

———. 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.

Greif, Avner, Paul Milgrom, and Barry R. Weingast. 1994. "Commitment, Coordination, and Enforcement: The Case of the Merchant Guilds," 102 *Journal of Political Economy* 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," 6 *Journal of Law, Economics, & Organization* 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*.

Kaplanoff, Mark D. 1991. "The Federal Convention and the Constitution," in Jack P. Greene and J. R. Pole, eds., *The Blackwell Encyclopedia of the American Revolution.* Cambridge: Basil Blackwell, Inc.

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," 27 *Journal of Economic Theory* 253–79.

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

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, CT: 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.

Meinig, D.W. 1993. *The Shaping of America: A Geographic Perspective on 500 Years of History.* Vol. 2: *Continental America, 1800–1867*. New Haven, CT: 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," 2 *Economics and Politics* 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," 48 *World Politics* 50–81.

Moore, Glover. 1953. *The Missouri Controversy: 1819–1821.* Levington: University of Kentucky Press.

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

Morrow, James D. 1994. *Game Theory for Political Scientists.* Princeton, NJ: Princeton University Press.

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," 45 *World Politics* 99–126.

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

Ordeshook, Peter C., and Olga Shvetsova. 1997. "Federalism and Constitutional Design," 8 *Journal of Democracy* 27–42.

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

Persson, Torsten, and Guido Tabellini. 1996a. "Federal Fiscal Constitutions: Risk Sharing and Moral Hazard," 64 *Econometrica* 623–46.

———. 1996b. "Federal Fiscal Constitutions: Risk Sharing and Redistribution," 104 *Journal of Political Economy* 979–1009.

Poterba, James, and Jürgen von Hagen, eds. 1999. *Fiscal Institutions and Fiscal Performance* Chicago: 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.

———. 1987. "The Lessons of 1787," 55 *Public Choice* 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 Arthur Meier Schlesinger, ed., *New Viewpoints in American History*. New York: Macmillan.

Solinger, Dorothy. 1991. "The Floating Population as a Form of Civil Society," paper for the 43rd 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, MA: Harvard University Press.

Stepan, Al. 1998. *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," 64 *Journal of Political Economy*, 416–24.

Tirole, Jean. 1988. *The Theory of Industrial Organization.* Cambridge, MA: MIT 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. 1999a. "Corruption and Federalism," working paper. UCLA.

Treisman, Daniel. 1999b. "Political Decentralization and Economic Reform: A Game Theoretic Analysis," 43 *American Journal of Political Science* 488–517.

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," 11 *Journal of Law, Economics, & Organization* 1–31.

Weingast, Barry R. 1997. "The Political Foundations of Democracy and the Rule of Law," 91 *American Political Science Review* 245–63.

———. 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, eds., *Analytic Narratives*. Princeton, NJ: 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* 691–715.

Wood, Gordon. 1969. *The Creation of the American Republic, 1776–1787*. New York: Norton.

———. 1991. *Radicalism and the American Revolution*. New York: Vintage Books.