Shirking and Stability in Federal Unions

Jenna Bednar¹
Department of Political Science
University of Michigan
jbednar@umich.edu

April 2001

¹Over several years I have benefitted from the suggestions of many colleagues. At the dissertation stage, I thank John Ferejohn, Brian Gaines, Avner Greif, Jack Rakove, Doug Rivers, Matt Spitzer, and Barry Weingast. For more recent suggestions I thank Sven Feldmann, Barry Friedman, Simon Hug, Ken Kollman, and Scott Page, and seminar participants at Stanford, Caltech, UC-Irvine, SUNY Stony Brook, the University of Pittsburgh, Iowa, the Claremont Graduate School, and the University of Minnesota. I thank the Olin Foundation for research support and the warm hospitality of the University of Southern California Law Center. This paper extends an unpublished work, "Federalism: Unstable By Design," which I presented at the conference "Law and Economics of Federalism," co-sponsored by the University of Minnesota Law School and Federal Reserve Bank of Minneapolis, May 1997, and at the 1999 meeting of the American Political Science Association, Atlanta, GA.
Abstract

Intergovernmental regimes, from treaty organizations to federations, never operate in perfect harmony. Our instinct is to fine-tune the institutions that regulate relations between the governments until we eliminate all tension. With a model of incomplete information, this paper explains the persistence of shirking in intergovernmental organizations by showing that except under special circumstances, no full contribution equilibrium exists. Instead, the optimal possible equilibrium includes a suboptimal productive level from the union with occasional serious disputes. The institutionalized shirking and periodic disputes can be reinterpreted as evidence of an institutional structure that works well, rather than needing repair.

The results are extended in two ways. First, the players can opt out of the union. In the symmetric example, exit options only increase benefits if they are superior to the union’s offerings: moderate to mediocre exit options reduce utility. In a second extension, the paper considers the effect of various forms of asymmetry on the equilibrium contributions by members. Here, we see that exit options may be of benefit to certain players, but not all.

The paper concludes with the suggestion that we design institutions that accommodate the natural tendency to shirk and efficiently manage any resulting tension. Although the model is motivated by federalism, the results may be interpreted generally for all repeated public good provision problems with a continuous action space and imperfect monitoring.
1 Introduction

In federalism we observe no perfectly harmonious unions; instead, even the most stable—the United States since 1865, Switzerland—are characterized by near-constant quibbling and, periodically, more serious disputes. Others, such as Canada, seem to be perennially at the brink of rupture. We would like to explain the persistence of shirking and the somewhat surprising stability of these intergovernmental organizations despite the intergovernmental rivalry.

Generally, we explain rivalry in terms of conflicting interests, for example, in federations, we explain interregional tension by pointing to the cultural or economic differences between the constituencies of the rival units.1 We know, however, that federalism is prescribed precisely when regions represent clusters of heterogeneous preferences (Tiebout 1956, Oates 1972, Ostrom 1991, Peterson 1995, Tullock 1969).2 Therefore, an excuse for intergovernmental tension that relies upon heterogeneity begs the question: if we implement federalism to satisfy regional differences, then how do we know when the diversity passes from necessitating federation to rupturing it? In other words, why does the ethnic rivalry between the Muslim and the Croats render Bosnian federation untenable, while the differences between French and English Canadians, or Czech and Slovak Europeans, can be settled peaceably (albeit not always together), and French and German Swiss manage to govern themselves harmoniously? Furthermore, if regional diversity explains federal instability, why did it take 75 years to endanger the United States and roughly 100 years to threaten Canada?3 Diversity cannot be the litmus test to predict federal failure, since it also seems to be the reason for the establishment of the federation, and it is present in thriving federations. At least, diversity is not a sufficient condition for instability, and it might not be necessary.

Others have proposed that citizen loyalty can be the glue that holds a federal union together. William Riker (1964, especially p. 111) suggests that a transference of loyalty from region to union might be the key to federal stability. Presumably, if all citizens want the union to work, they will not pursue actions that jeopardize it. When citizens do not believe in the federation, the federation

---

1See, for example, Franck 1968, Friedrich 1968, Hicks 1978, Lemo 1991.
2See also Cremer & Palfrey (1996, 1997) whose formal models indicate that decentralization is most attractive to populations with divergent preferences.
3For discussions of the timing of federal instability, see Weingast (1994) and Bednar, Eskridge, and Ferejohn (1995).
will fail. Riker's argument lies in good company (Elazar 1987, Franck 1968, Beer 1993), but the argument's reasonableness dissipates when we press for a micromotive foundation: we don't know what initiates the commitment to the union, nor do we have an idea of what causes the interest in the union to disintegrate.

However intuitively appealing arguments about diversity or loyalty transference might be, the theories' strengths are compromised by their idiosyncratic nature; they treat failing federations on a case-by-case basis, diagnosing problems as they arise. Typical of the literature is the voice of one comparative theorist, K. Hicks (1978, 171): “In assessing causes of failure or success [of federalism] it is seldom a matter of identifying a unique factor; it is rather a question of judging which elements in a complex situation were most responsible for the result.” Without a doubt, this approach has provided excellent explanations for troubles plaguing a particular federation at a specific time. Prescriptions for stability follow logically from this ad hoc approach, unfortunately treating symptoms rather than improving our understanding of the disease.⁴

When we see intergovernmental rivalry, we are tempted to blame the institutions that regulate interaction within the union. We want to adjust the institutions to eliminate all rivalry, leaving us with a perfectly harmonious union. This paper turns around the objective, to ask: since no “perfect” federation exists, that is, all federations are to some degree “unstable” because all exhibit intergovernmental tension, what causes many federations to survive, despite the tension? In fact, this paper argues that not only is tension compatible with stability, it is necessary for it. In so

⁴The formal comparative politics literature (in particular, scholars of contemporary Russia) features recent articles advancing our understanding of the way that institutions affect the outcomes attained in federal systems. See, for example, Aranson (1995), Chen and Ordeshook (1994), Ordeshook (1995, 1996), Ordeshook and Svetsova (1995), Solnick (1998), Tierman (1999), Weingast (1997), and de Figueiredo and Weingast (1997). However, with the possible exception of Tiersman, in none of these articles do the authors provide a theory for persistent opportunism in federal systems; the positive literature assumes that governmental agents will take advantage of one another whenever possible, without probing its underlying causes or even addressing whether or not instability is inevitable. Most of the formal work on federalism is concerned with problems of redistribution or overcentralization, both problems that can lead to instability, but with the exception of de Figueiredo and Weingast (1997) and Tiersman (1999), the formal literature does not explain how the governmental structure induces noncooperative behavior and generation of externalities. Tiersman embeds a bargaining model into a federal structure where the center has the power to punish errant regions. In his model, the center and the regions have different levels of strength; headstrong regions are able to capture rents from a weak center. The bargaining model is especially relevant for Russia, where federalism is constituted by a series of bilateral agreements between the center and each province. The form may be gaining popularity, as the European Union negotiates a la carte memberships including opt in and opt out clauses, but most federations do not have this structure. I will discuss the de Figueiredo and Weingast paper at greater length below; the equilibrium in their model does not capture continued opportunistic behavior.
doing, this paper is able to propose a contrary interpretation of rivalry: rather than serving as an indication that the federation is in trouble, it may mean that the stabilizing institutions are working just right.

My chief interest in this paper will be to explain the persistence of intergovernmental rivalry—even in a stable federation like the United States—by understanding its source. Intergovernmental tension and federal instability are closely related: failed federations show symptoms of tension before the system falls apart. Indeed, one puzzle remaining to the federalism literature is why regional and central governments would engage in tension-causing opportunistic behavior when it would seem to risk failure of the union. The question remains open: is there something about federalism that creates a systemic bias toward weakness? Without a theory of federalism that recognizes the inherent instability of the system, we are unable to suggest durable institutional remedies to support federalism in the long run. The aim of this paper is to contribute to understanding the pathology of federal instability, to understand the causes and consequences of intergovernmental federal tension. With a better understanding of the causes of intergovernmental tension and whether or not it can be fully eliminated, we are better prepared to predict how a federation will react to constitutional prescriptions.\(^5\)

This paper pursues a general theory of intergovernmental tension, distilling characteristics common to all federations. While pre-existing animosity may aggravate federal tension, the model here highlights a different peril: the decentralized structure, the very thing that makes them federal. Therefore, all federal systems are susceptible to destabilizing intergovernmental competition. In other words, federalism is unstable by design. Intergovernmental tension is inherent. Second, while federations all contain the seeds for instability, they are not doomed to failure. Federal stability is not about the elimination of federal tension, but instead, about managing destructive behavior. I argue that federations may survive with moderate tension, as the quite stable modern United States has managed. The model indicates how a primitive institution is able to maintain stability despite being unable to prevent opportunism, and in the discussion, I suggest that more sophisticated

\(^5\)Ideally, we’d know more about the process and development of federal instability as well, that is, the ebb and flow of cooperative behavior. Equilibrium analysis, the standard formal modeling form, and the one employed here, is constrained in its ability to inform us about dynamic development. However, equilibrium analysis can help to build an understanding of the source and effect of intergovernmental tension.
institutions may more efficiently bolster the union.

I make two points about the possibility of stability. First, the federal mechanism's design tolerates some disappointing outcomes, and therefore federalism can recover from moderate spells of destabilizing behavior—competitive, externality-generating action—by its member governments. Second, the theoretical results suggest a functional theory of the role of institutions in providing stability. I outline these results here, concluding that it is critical to pay attention to the functional link between the cause of destabilizing intergovernmental tension and the cure offered by the institution.

As I extend the model, I show how the utility the members get from the union depends not just on the internal efficiency of the union—the subject of most of the fiscal federalism literature—but also upon the participants' outside options. Stability and the union's productive capacity are both affected by exit options. I show how mediocre exit options—ones that the participants would not choose over membership in the union under normal circumstances—affect the productivity of the union when truly mediocre, and as the exit option improves, but remains below the offerings of a productive union, either no union is possible, or it will break apart at the first bit of bad luck.

Further extensions discuss several types of asymmetry: in utility from the union, in contribution to the union, and in likelihood that non-compliance triggers a reaction in other participants, among other modifications. While asymmetry does not change the result that federations operate below peak efficiency, it does show that power asymmetries have a great effect on the character of the union. The section suggest ways that the general model could be tailored for analysis of specific federations.

Several working assumptions will govern our generalization of federal interaction.⁶

(A1) First, I assume that the governments—regional and central—are independent: they have authority to make meaningful decisions, which they make on their own. In order to separate incentives from institutional power, I make no assumption about the ability of any government to

⁶Together, these assumptions define the essential components of a federal system. Bear in mind, however, that federalism is best considered as placement of a polity on a continuum of relative decentralization. The only quick marker that defines a federation from a unitary state is that in a federation, true power-sharing exists between levels. Unitary states may devolve authority for administrative convenience, but that authority may be recalled by the center at its own pleasure. Federalism implies constitutional recognition of jurisdictional separation.
coerce another.\footnote{While it is tempting to think of the center as having much more power than any regional government, this impression is based upon the institutional arrangement of our modern American model. For example, de Figueiredo and Weingast (1997) assume a monitoring and corrective power for the center. This paper separates agent motivation and the institutional environment; it is possible that the center could be held captive by dominating regional governments, as we know from the historical experience with United States under the Articles of Confederation, or Yugoslavia in this decade, or any other federation where one or more regions have the ability to stymie the central government, including, arguably, Canada. Therefore, in order to keep the theory as general as possible, I avoid any power assumptions. However, Corollaries 1–3 in Section 5 consider the effect of an induced asymmetric power relationship, where one government’s utility from the union is different from the others.}

(A2) Second, each government chooses a level of compliance with the constitutional rules that govern its interaction with the other federal participants. It will be most convenient to think of the level of compliance as a percentage, ranging between zero and one, where one equals full compliance.

Zero-percent compliance is not equivalent to withdrawal. If a government remains in the union, but does not comply with its terms, it fails to make any of the effort or sacrifice expected of it, as a participant, but it still hopes to gain the benefit of membership; it behaves as a free rider. With withdrawal, on the other hand, a government loses nothing, but also stands to gain nothing; it is entirely disassociated from its former compatriots. To understand fully the distinction between zero-percent compliance and withdrawal, consider the options available to the central government. Zero-percent compliance implies that the center makes no sacrifice to continue the union; such a union would be characterized by extreme centralization at the center’s convenience. We can think of periods in the recent history of India, Mexico, or the Soviet Union for examples of highly centralized federations, where the center would approach near-zero levels of compliance. Withdrawal by the center, on the other hand, means a dissolution of the federal union through common consent, such as New Zealand in 1876. Section 5 considers effect of exit options on the stability of the union.

While individual levels of compliance are not observable by other agents, in Section 6 I relax this assumption somewhat by introducing asymmetries in the likelihood that an agent’s action triggers the punishment regime.

(A3) Third, I assume that non-compliance is detrimental to the union. Common words for non-compliance include shirking, encroachment, burden-shifting, and favoritism. All federations are familiar with these symptoms.

Regions (the sub-level of government, such as states or provinces) try to shift the burdens of fed-
eral membership onto the backs of other regions, by arguing over tax incidence, trade arrangements, location of industry, redistribution, and the like.⁸ Political economists often focus on competition between regions (Maggi 1993, Persson & Tabellini 1996a, 1996b, Cremer & Palfrey 1996, 1997, Caplan 1996, Weingast 1994, 1995, Bendor & Mookherjee 1987), and therefore tend to overlook the contribution of the central government to instability. The central government is not an angelic constraint;⁹ it too is guilty of contributing to the risk associated with participation in a federation. Some of the political burden-shifting battles seep into the central government’s agenda: charges of favoritism and bias disturb a federal balance as much as battles between regions.¹⁰ Centralization or peripheralization, when carried to extremes, threatens federalism as much as burden-shifting.¹¹

The paper condenses the many forms of opportunism simply to be deviance from prescribed behavior, or more briefly, non-compliance. The model will uncover the motivation for shifting distrust to “defensive” non-compliance and it will explain the persistence of moderate non-compliance in otherwise stable federations, a puzzle as yet unexplained in the literature.

Therefore, non-compliance stimulates intergovernmental tension, which can manifest itself as bickering or, more dramatically, as internal power struggles. Intergovernmental tension may also lead to an agent’s reevaluation of the benefits of remaining in the union, and push an agent to

---

⁸Perhaps guided by Madison, who suspected that the tendency to federal instability was “rather to anarchy among the members, than to tyranny in the head.” (Fed. 18) a healthy portion of the federalism literature is consumed with interregional rivalry of this sort. The earliest, and still most active, formal research on federalism is labelled “fiscal federalism,” where economists draw upon Tiebout models to generate optimal taxation and distribution schemes. Some of this work is becoming more political: for example, Aranson (1995) contributes to the literature on federal stability by building from the Tiebout notion of the mobile citizen. He assumes that the center wants to expropriate rents from the regions and that the regions collectively can constrain the center, provided that they are able to overcome the collective action problem. The paper suggests that there may be an optimal number of districts in a federation.

⁹For indications of its usefulness, however, see Wechsler (1954), Weingast (1995), Montinola, Qian, and Weingast (1995).

¹⁰With some clever tweaks to the interregional rivalry models, to allow for capture at the central level by regional agents, the standard model of federalism, based upon diverse preferences between regions, can be maintained (Dixit & Londregan 1996). Bednar (1998) and Voklen (1999) motivate central governmental encroachment, the former in a decision-theoretic model, and the latter, game-theoretically.

¹¹See, for example, Riker’s (1964) judgment that over-peripheralized federations cannot survive, and Hamilton’s similar warning, with the words dividere et imperium to convince by fear the unconvinced of federalism’s virtues. (See especially Federalist No. 7.) Bednar et al (1995) argue that the mere threat of centralization can destabilize a union; see the section on Canada.
withdrawal.

(A4) To simplify the portrayal of institutional incentive schemes, I include a punishment regime as part of the strategy of participating players, something akin to a war of attrition, or universal free-ridership. The punishment regime demands significant cost from all players and is never preferable to compliance in a single period. In a punishment regime, all governments quit their cooperative, compliant behavior, in effect punishing everyone else. Three motivations exist for punishing everyone:

1. First, it is possible that the players cannot observe directly any other player's level of compliance. We would expect this to be true most often in politics where there is little reliable information exchanged between governments. However, this assumption also applies to advanced societies, such as the United States. Here, instead of having too little information, we have too much, again making effective monitoring of others unreasonable. Finally, some situations are just too complicated to determine who is responsible for disappointing outcomes: consider the environmental degradation of the Gulf of Mexico. With any one of perhaps twenty states a contributor to the pollution around the Mississippi delta through the extensive river system draining to the Gulf, which one should we single out to punish?

2. It is often very difficult to selectively punish governments. In the United States, with the notable exceptions of California, Texas, Alaska, and Hawaii, population centers tend to be situated near state borders. Therefore even if targeted measures such as fines were politically feasible, they might not be effective; citizens or industries can easily dodge them by crossing over into a neighboring state.

3. Finally, it is important to consider the punishment in the context of the equilibrium (derived in section 4). In equilibrium, no one “cheats”, but all deviate from full compliance by some moderate amount. Therefore, if the federation does not return the benefits that the players expected, each participant would examine the actions taken by others, in an attempt to decide

---

\textsuperscript{12}It is possible to consider a hierarchy of punishment regimes available to the federal participants, with the war of attrition being the least costly, but also, perhaps, the least efficient. Next in line would be domestic conflict, where the participants' private effort level is depleted as well. This model does not embrace civil war specifically as the utility function is too general, but it could be easily adapted to do so.
who was responsible, and see that all players were deviating slightly. Again, the questions become: whom do you punish, and when?

(A5) Fifth, I assume that each government wants to maximize the utility it provides to its own constituents, represented simply as each government’s utility. If it participates in the union, the utility is a function of the benefits of the union and the costs and sacrifices associated with participation. If it withdraws, its utility is exogenously determined and commonly known.

(A6) Lastly, I assume that there is some cost to compliance.

To explain why any government in a federation would fail to comply with the terms of the union, we might be tempted to stop with the last assumption: if compliance is costly, then of course participants would want to shirk. However, this argument ignores (A3-A5). While non-compliance is tempting, union participants might not do it just as most of us would not shoplift. We would—most of us—prefer that the shop continue to exist and especially that we continue to be at liberty to shop there, more than we would prefer to have something (especially something pocket-sized) for nothing. We don’t shoplift because we fear the consequences more than we value the benefit of having both goods and money.

So again, we ask, why do governments ever shirk? Well, in some idealized conditions, they wouldn’t, as I discuss in section 3, (after setting up the model in section 2). Section 4 explores the more general case: when we relax the strict assumptions made in Section 3 to better fit political reality, we see that shirking is a natural part of federalism’s operation. Section 5 adds exit options and Section 6 shows how the model can be adapted to generate testable hypotheses and discusses the effect on federalism of various forms of asymmetry, providing some insight for the design of effective institutional intervention. Section 7 discusses the results and concludes.

2 The Model

At the start of each period \( t \), all eligible governments, \( i, i \in \{1, \ldots, N\} \), (A1) choose to participate in the union or to withdraw (A2). I suppress \( t \) when doing so causes no confusion. A choice of withdrawal is permanent. If a player exits, let \( W \) define the present discounted value of seceding (a government’s “exit option”), including any expected penalty from the act of secession. With this
we define a participation constraint.

If a government chooses to participate in the union, a constitution prescribes its behavior. Actions, such as passing legislation, taxation, or implementing policy, can be considered as a degree of compliance with the constitutional rules. Therefore players select some percentage of compliance: formally, in each period $t$, each player $i$ chooses a level of compliance $a_{it}, a \in [0,1]$, with $a_{it} = 1$ representing full compliance with the constitution (A2). Because the agents are not limited to a binary choice of comply or shirk, but may choose a percentage of compliance in a continuous choice space, this model can capture subtle phenomena such as moderate slippage from full cooperation. Notice the fundamental difference between occasional full non-compliance and continuous but minor deviations from full compliance: this paper will explain the latter, a phenomenon that occurs in all federations; the model will include the former as a punishment mechanism to increase the overall level of compliance.

Stochastic uncertainty is captured by a continuous and nonatomic random variable, $\omega$, with an expected value of zero $E(\omega) = 0$. Since the stochastic term can assume positive or negative values, the model can capture unexpected benefits as well as unexpected losses.\footnote{Compare this model to the decentralized decision-making model of Bendor and Mookherjee (1987), where uncertainty enters as a probability that a player's effort will not successfully translate into a contribution. In the present model, uncertainty can assume positive or negative values. For the purpose of understanding interaction in a federal system the latter is preferable since it allows for good random shocks as well as bad; an errant member might get lucky and do something that is unexpectedly beneficial for the federation. Also, the Bendor and Mookherjee model assumes a binary choice of action: work hard or slough off, while the action space in the present model is continuous and infinite.} Each $\omega$ is a policy shock that governments do not observe directly.

The stage game utility to each participating government is a function of its own action, the level of compliance of the other players, and the noise term: $U = U(a_i, a_{-i}, \omega)$. Federalism suffers from the classic have-your-cake-and-eat-it-too syndrome of a public good provision problem; captured with the assumption that government utility decreases in its own level of compliance, $\frac{\partial U}{\partial a_i} < 0, \forall i$ (A6), but increases in the level of compliance of all other players, $\frac{\partial U}{\partial a_j} > 0, \forall i, \forall j \neq i$ (A3). While actions are not directly observable, they can be inferred, albeit imperfectly, by the performance of the union as reflected in each player's utility $U(\cdot)$.

Finally, strategies describe a player's action for all possible histories of utility; each stage maps
the utility histories into a continuous choice space of action, \( s^*_t : H^t \to A^t \), so that actions are a function of history, \( a^t(h^t), a^t_i \in [0, 1] \). At the start of a period, each player chooses an action, the actions are aggregated in a manner that defines the union’s productivity, and the productivity is distorted by idiosyncratic noise. Payoffs to each agent depend upon the value of the union good and their own action. The objective of each agent is to maximize expected discounted value of utility. Government impatience is represented by a non-zero discount rate, \( \delta \in (0, 1) \). Governments are risk neutral and maximize \( E[\sum_{t=0}^{\infty} \delta^t U(h^t)] \) (A5).

A Nash equilibrium is a strategy profile \((s^*_1, \ldots, s^*_n)\) which satisfies

\[
E_{s^*_1, \ldots, s^*_n} \left[ \sum_{t=0}^{\infty} \delta^t U \left( s^*_t(h^0, \ldots, h^{t-1}), s_{-i,t}(h^t), \omega_t \right) \right] 
\geq E_{s^*_1, \ldots, s^*_n} \left[ \sum_{t=0}^{\infty} \delta^t U \left( s_{it}(h^0, \ldots, h^{t-1}), s_{-it}(h^t), \omega_t \right) \right]
\]

for all governments and feasible strategies \( s_i \).

### 2.1 Trigger Mechanism

Although the players cannot monitor the levels of compliance of one another, they can infer it, subject to uncertainty, by evaluating the single-period utility that they each receive from participating in the union. All federations employ some institutional mechanism to induce cooperation. Here, I represent these institutions simply as a trigger mechanism (A4). If the utility falls below a threshold, \( \tau \), players refrain from cooperating for a finite number of periods.\(^{14}\) Therefore, the utility maximizing equilibrium takes the following form:\(^{15}\)

\[
a_{i,1} = 1 - \theta \\
a_{i,t+1} = 1 - \theta \quad \text{if} \quad U_{i,t}(F(\vec{a})) > \tau
\]

\(^{14}\)It is important to keep in mind that the punishment regime is not the same as withdrawal, or secession from the union. Players still “participate,” but they do not cooperate with one another by complying with the constitution.

\(^{15}\)Abreu, Pierce, and Stacchetti (1990) show that efficient equilibria rely on grim trigger strategies. While their result is proven for a finite action space, and so does not translate directly to this paper’s infinite action space, the intuition is equivalent. De Figueiredo and Wengast (1997) develop a model with a binary choice space and imperfect monitoring and, in keeping with the Folk Theorem literature, support full compliance in equilibrium. Like Triesman (1999), their model differs significantly from the present model in combining the incentives of the governments and the institutional structure that supports it, by assuming that the center has monitoring and punishment powers over the regional governments. While neither approach can disentangle federal instability from institutional weakness, their models have several advantages, including intuitive appeal, and are appropriate for many federations.
\[ a_{i,t+j} = 0 \quad \text{else, for} \quad j = 1, \ldots, T \quad \text{periods} \]

where \( \theta \in [0, 1] \). Let \( X_i \) represent the utility received in each period of the punishment regime and \( F \) be a function that determines the productivity of the union, where \( \frac{\partial F}{\partial a_i} = \frac{\partial F}{\partial a_j} \forall i \) and \( j \). For example, the function \( F \) could be equivalent to the sum of the contributions plus the noise term: \( \sum a_i + \omega \). I assume that \( U_i(F(\bar{a})) > X_i \) for any \( a_i \) (A4).

The paper is interested in instances where \( \theta > 0 \), indicating partial lack of compliance, activity that has the potential of causing intergovernmental tension.

### 2.2 The Value Function

Governments maximize the present discounted value of being in the union, wanting to make their membership as worthwhile as possible. Note that a player compares \( V_i \), its expected value from participation, to \( W_i \), its outside option. To solve for \( V_i \), I construct the following Markov dynamic programming problem. Let \( a^* \) be a candidate for an equilibrium. The present discounted value for agent \( i \) if it plays a strategy of \( a_i \) in each period and if the others play \( a^* \) is defined as follows:

\[
V_i(a_i|a^*_{-i}) = U_i(F(a_i, a^*_{-i}, \omega)) + (1 - p)\delta V_i(a_i|a^*_{-i}) + p\left(\Delta^T X + \delta^{T+1}V(a_i|a^*_{-i})\right) \quad (2)
\]

where \( \Delta^T \) is the sum of all discount parameters to the \( T \)th period and \( p = \text{prob}(F(a, \omega) < \tau) \). The value function captures both the costs and benefits of shirking (the temptation to contribute less than \( a_i = 1 \)). The first term on the right-hand side represents the marginal benefit of shirking, while together the second and third terms represent the marginal cost of shirking. The second term on the right-hand side represents the continuation value to participant \( i \) of successful coordination (when the total value to the union is above \( \tau \)), and the third term the remaining cases, when the value falls below \( \tau \) and the union enters a punishment regime. Letting \( V \) and \( U \) represent \( V_i(a_i, a_{-i}, \omega) \) and \( U_i(F(a_i, a_{-i}, \omega)) \), respectively, Eqn. 2 simplifies to:

\[
V = U + (1 - p)\delta V + p\Delta^T X + p\delta^{T+1}V
\]

\[
V - (1 - p)\delta V - p\delta^{T+1}V = U + p\Delta^T X
\]

\[
V \left(1 - \delta(1 - p) - p\delta^{T+1}\right) = U + p\Delta^T X
\]
\[ V = \frac{U + p \Delta T X}{1 - \delta(1 - p) - p \delta^{T+1}} \]

\[ V = \frac{U + p \Delta T X}{(1 - \delta) + p(\delta - \delta^{T+1})}. \]  

Eqn. 3 provides a generic value function for this class of problems.

The vector of strategies \( \alpha^* \) is an equilibrium if and only if

\[ V_i(a^*_i, a_{-i}) \geq V_i(a_i, a^*_{-i}) \quad \text{for all } a_i \in [0, 1]. \]

Equivalently, \( \frac{\partial V_i(a_i, a^*_{-i})}{\partial a_i} = 0 \) and \( \frac{\partial^2 V_i(a_i, a^*_{-i})}{\partial a_i^2} \leq 0 \), provided \( V_i \) is concave.

Each player selects its optimal strategy by making a calculation about what it expects to gain by deviating from full compliance (the first term of Eqn. 2). It compares this benefit of deviation to the cost of deviation, which is the increased probability that \( U \) falls below the threshold \( \tau \). Deviating offers short-term gain, but comes at a price of increasing the likelihood that the union falls out of the cooperative equilibrium and into an unproductive punishment regime. The agent deviates until the marginal benefit of deviation equals the marginal cost.

To find the value of \( a_i^* \) that maximizes the value function, we solve as follows. From Eqn. 3 we take the derivative w.r.t. \( a_i \):

\[ V = \frac{U + p \Delta T X}{(1 - \delta) + p(\delta - \delta^{T+1})} \]

\[ V' = \frac{\left((1 - \delta) + p(\delta - \delta^{T+1})\right) \left(\frac{\partial U}{\partial a_i} + \frac{\partial p}{\partial a_i} \Delta T X - (U + p \Delta T X) \left(\frac{\partial p}{\partial a_i} (\delta - \delta^{T+1})\right)\right)}{((1 - \delta) + p(\delta - \delta^{T+1}))^2} \]

An equilibrium satisfies \( V' = 0 \), so it suffices to drop the denominator:

\[ 0 = \frac{\partial U}{\partial a_i} \left((1 - \delta) + p(\delta - \delta^{T+1})\right) + \frac{\partial p}{\partial a_i} \Delta T X \left((1 - \delta) + p(\delta - \delta^{T+1})\right) + \cdots \]

\[ - \frac{\partial p}{\partial a_i} U(\delta - \delta^{T+1}) - \frac{\partial p}{\partial a_i} p \Delta T X (\delta - \delta^{T+1}) \]

\[ = \frac{\partial U}{\partial a_i} \left((1 - \delta) + p(\delta - \delta^{T+1})\right) + \frac{\partial p}{\partial a_i} \Delta T X (1 - \delta) - \frac{\partial p}{\partial a_i} (\delta - \delta^{T+1}) U \]

\[ = \frac{\partial U}{\partial a_i} \left((1 - \delta) + p(\delta - \delta^{T+1})\right) - \frac{\partial p}{\partial a} \left(U(\delta - \delta^{T+1}) - \Delta T X (1 - \delta)\right) \]

\[ 0 = \frac{\partial U}{\partial a_i} \left((1 - \delta) + p(\delta - \delta^{T+1})\right) - \frac{\partial p}{\partial a_i} \left((U - X)(\delta - \delta^{T+1})\right) \]

Eqn. 5 specifies a necessary condition for the level of compliance that maximizes the value function.
3 Conditions Necessary to Sustain Full Compliance

3.1 Complete Information

To derive intuition about strategies, we consider the full information case,\(^{16}\) where we suppress the random variable \(\omega\), so the value of the union plainly reveals whether or not any agent shirked. Each agent chooses an action \(a_i \in [0, 1]\). Each government's payoff is a function of the joint sacrifices made by all participating members and the government's attempt (if any) to cull more from the union for the benefit of its constituents, \((1 - a_i)\), making the single period utility function \(U = U(a_i, a_{-i})\), with (A3) and (A6) continuing to hold. Participation is still subject to the constraint: \(V > W\).

In this game, unlike the prisoner's dilemma, the action space is continuous, taking any value between 0 and 1, inclusive. In the prisoner's dilemma, the choice space is discrete: players may cooperate or defect, making the prisoner's dilemma unable to account for slippage from full cooperation. Also, this model includes more than two players. However, a comparison to the prisoner's dilemma provides intuition.

In a single-shot prisoner's dilemma, no player wants to be caught complying when the other deviates, but each would like to deviate if the other complies. Although all players prefer the cooperative equilibrium of (comply, comply) to (defect, defect), in the single shot prisoner's dilemma, the unique Nash equilibrium is for both players to deviate, yielding a sub-optimal payoff. Clearly, this model of federal interaction is not sufficient as it predicts the existence of no federations. More promising is a repeated model.

While many equilibria exist in the repeated game setting, we concentrate on the full compliance equilibrium, where \(a_i = 1\).\(^ {17}\) As long as the discount rate is sufficiently low (that is, the discount factor \(\delta\), must be sufficiently high), compliance can be sustained with a grim trigger punishment mechanism. Strategies are contingent upon the play in the previous round: each agent cooperates until one defects, at which point they pull the "grim trigger," and defect forever.

\(^{16}\)See also de Figueiredo and Weingast (1997) for an analysis of the complete information case with linear utility functions and a binary action space.

\(^{17}\)This result is similar in feel to the Folk Theorem. However, the Folk Theorem is generally applied only to finite action spaces.
**Proposition 1:** With complete information, full compliance can be achieved with a trigger strategy in which deviations are punished forever by setting the threshold, \( \tau \), equal to full contribution from each member.

\[
a_{i,t} = 1 \\
\quad a_{i,t+1} = 1 \text{ if } U_i(F(a_{i,t}, a_{-it})) > \tau \\
\quad \quad = 0 \quad \text{else.}
\]

**Proof:** It suffices to show that no player can benefit by deviating. The player compares its expected utility of complying with the expected utility from deviating, choosing the action that gives it the higher expected payoff. Without any information uncertainty, any deviation is revealed. Therefore, the threshold, \( \tau \), is set such that it tolerates no deviation from full compliance. For simplicity, let the vector \( \bar{a} \) represent the symmetric action taken by all participants: i.e. \( \bar{a} = 1 \) represents full compliance by all. By assumption (A4), \( U_i(F(\bar{a} = 1)) > U_i(F(\bar{a} = 0)) \) and \( U_i(a) > X \), for all \( a \), so no player would prefer to deviate, knowing that to do so would certainly trigger the punishment regime, for a sufficiently high discount rate.\(^{18}\)

While the grim trigger strategy is theoretically efficient, giving the most bang for buck in terms of ensuring compliance, it is severe and unforgiving. It is also irrelevant from a practical perspective: no player would choose to remain forever in a non-complying, non-productive union, when it could withdraw from the union and wash its hands of the matter. In practice, we would expect that the grim trigger strategy would become a strategy of union dissolution.

Another strategy option exists to the agents: full compliance can also be sustained with a finite punishment strategy, where players respond to less-than-full compliance with non-compliance (as opposed to withdrawal), for a specified number of periods, rather than forever, as in the case with the grim trigger.

\(^{18}\)See the proof for Proposition 2 in the Appendix.
Proposition 2: The full compliance equilibrium can be sustained with a finite punishment regime, defined as follows:

\[
\begin{align*}
    a_{i,t} &= 1 \\
    a_{i,t+1} &= 1 \quad \text{if } U_{it}(F(a_i, a_{-i})) > \tau \\
    a_{i,t+j} &= 0 \quad \text{else, for } j = 1, \ldots, T \text{ periods.}
\end{align*}
\]

PROOF: See appendix.

The proof generates the following comparative statics: \( T \), the minimum number of punishment periods, will vary inversely with \( U_1(F(\bar{a} = 1)) \) and \( \delta \), the benefit of participation and the discount factor, respectively.\(^{19}\) Therefore, as the players become increasingly patient, the minimal number of punishment periods necessary to sustain full compliance declines. Furthermore, as the union becomes increasingly efficient (for different functions \( F \)), giving the players greater returns to their effort (compliance), they are increasingly willing to cooperate, so fewer punishment periods are needed to sustain compliance. These results continue to hold in the more general cases below.

If a grim trigger strategy is in practice equivalent to permanent withdrawal, the finite punishment regime is the strategy that makes the union’s regeneration possible. In the finite punishment regime, no member contributes, so the union is not productive, but the nominal association persists, ready to be revived at a later date. It is the most primitive of institutional mechanisms available for coercing all participants to comply with the terms of the federal bargain. Implementing the punishment mechanism sacrifices the benefits of cooperation for some time, but it does not entail further costs, such as dipping into the reserved benefit \( (a_i) \). The next most primitive method to induce compliance is domestic conflict, a means that does require personal sacrifice beyond the loss of opportunity. These measures are implemented after undesirable outcomes are felt, but in equilibrium they operate as ex ante deterrents, and are only used when environmental circumstances generate unfavorable levels of union productivity.

\(^{19}\text{See Eqs 11 & 13 in the Appendix.}\)
3.2 Incomplete Information and Linear Cost and Benefit curves

The section above demonstrates that with full information, a full compliance equilibrium can be sustained with a trigger strategy. However, even with incomplete information it is possible to find conditions that support full compliance in equilibrium, where $\theta = 0$, but these conditions demand strong assumptions about the shape of the cost and benefit curves, making full compliance highly improbable.

Proposition 3: If the probability of punishment and the benefit of deviating are linear in the level of compliance, then any symmetric stable equilibrium requires either full compliance or no compliance. No stable shirking equilibria exist.

PROOF: For a formal proof, see appendix. The governments comply at the level where the marginal cost of deviating equals the marginal benefit of deviation. If the cost and benefit curves are linear in the amount of compliance, then the marginal cost and benefit curves are level at a constant, giving us just three possibilities: (1) the marginal cost is always greater than the marginal benefit, for all levels of compliance, (2) the marginal cost is equal to the marginal benefit, or (3) the marginal cost is less than the marginal benefit of deviation. If the first is true, then no government will be tempted to deviate from full compliance. If the second is true, while equilibria with slippage may exist, they will not be stable. In the final case, governments will deviate fully, and the federation cannot be sustained.

In order for Proposition 3 to be satisfied, we must assume that the probability of punishment and the benefit of deviating are linearly related to the level of compliance. Above I indicated that these were strong assumptions, and therefore that the conditions were unlikely to be satisfied. We should satisfy ourselves that these conditions are indeed unlikely by considering further what is meant, exactly, by linearity. Perhaps the best way to talk about the unrealistic limitations of these conditions is by discussing what they exclude.

And, since the severity of the punishment is manipulable by the agents, they will probably adjust the marginal cost curve downward so that it just surpasses the marginal benefit curve.
In order for the conditions specified in these propositions to hold, none of the following may be true:

1. The additional responsibilities of autonomy become increasingly costly at the margin, making each additional dimension of autonomy less desirable than the one preceding it.

2. Citizens are indifferent about which level of government provides policy for moderate ranges of shared power, but dislike overly centralized or overly peripheralized federations. (They believe in the federal system.)

3. The likelihood of being caught is small at low levels of deviation, but for higher levels, the transgression becomes blatant, and the risk of punishment accelerates.

I will consider each in turn.

In a federation, regions compete for wealthy citizens, industries, and federal outlays. Citizens, companies, and federal money have high marginal return as the first few citizens or dollars are added, but in each case, the marginal benefit decreases as the numbers grow. Additional citizens and industries make further demands on the regional government, and money from the federal government is often tied to additional regional responsibilities, be they matching funds or installation and administration of additional programs. The increase in responsibility begins to weigh in against the initial attractiveness of governmental transfers, as the first condition implies.

Likewise, a central government can face decreasing returns to the centralization of control. Standards such as rail gauge, traffic signals, and electrical currents have immediate benefits; community library book purchases and training of police are possible to centralize, but the marginal benefit is slim, at best.

In the second condition above, citizens are tolerant of their government’s rivalry up to a point and reward their agents accordingly, but become increasingly suspicious of further claims to power.  

Consider the division of governmental authorities to be like a constant sum game, where the central and regional governments share a fixed, finite number of responsibilities. Citizens might be

\footnote{To return to Beer and Elazar, if not the loyalty argument of Riker, it may be the voters’ belief in federalism and decent governance that triggers the shift from praise to punishment as the politicians become more power-greedy.}
indifferent about the location of authority for moderate sharing of power, but dislike the extremes of total coordination or complete local autonomy.

Two examples from Canadian politics provide intuition. In the Canadian province of Quebec, the sovereigntist Parti québécois and Bloc québécois face serious opposition from the more moderate Liberal Party, and the voters have twice rejected a referendum proposal to secede from the Canadian union. Although the Québécois appreciate their regional agent’s efforts to gain more autonomy,22 the majority of voters are yet to provide a mandate for a purely selfish provincial agent. The Québécois value Canadian union positively, with a harmonious balance between central and regional authorities. They prefer moderation to extremism.

In a similar vein, Canadians grew suspicious of their Prime Minister, Pierre Trudeau, when he initiated an addendum to the Constitution to protect individual and community rights, over the protests of some provinces. While legally the federal government had the right to ignore the provinces in constitutional amendments, (a power the Supreme Court affirmed), political pressure from citizens across the country eventually persuaded Trudeau to back down and compromise with the provinces.

These anecdotes can be translated to condition 2. Citizens would reward agents who tug the system within a moderate range, but reject agents who try to overcentralize or peripheralize the federal distribution of power. Cavalier attitudes like Trudeau’s, displaying disrespect for the federal structure, are punished by the threat of being voted out of office. The marginal political benefits of deviance decrease with higher level of shirking.

Conditions 1 and 2 generate a concave benefit function, condition 3, a non-linear cost function. Recall that the cost of deviance is a calculation of the probability of getting caught (when the value of the union falls below the threshold) multiplied by the loss incurred when the federation enters a non-cooperative punishment regime. Because of environmental uncertainty, the threshold to trigger the punishment regime contains a little flexibility. For high levels of compliance, shirking a small amount is unlikely to trigger the punishment regime. Equivalently, when governments are

---

22In fact, students of Canadian politics have noticed a curious empirical regularity that voters often elect different parties to the provincial and federal legislatures. One popular explanation is that the citizens are balancing the powers.
hardly complying ($\bar{a} \approx 0$), the marginal effect of increasing the level of compliance is likely to have a negligible effect on the probability of punishment, as the federation is almost certain to enter the punishment regime anyway. However, for some intermediate levels of compliance, it is likely that even small deviations can have a large effect on the probability of punishment, acting as the proverbial straw that broke the camel’s back.

The presence of any one of the above conditions is sufficient to nullify Proposition 3 because it introduces non-linearity in the cost or benefit of compliance. Agents will achieve positive value from limited shirking. The level of non-compliance, $\theta$, becomes positive and $a^* < 1$. When any one of these conditions holds, full compliance with the terms of the federal bargain cannot be expected.

4 Modeling Inherent Opportunism

It is clear that in order to rule out full compliance, but to support some contribution in equilibrium, at least one curve—cost or benefit—must be non-linear, and the marginal benefits and marginal costs of shirking must be equal within the range of $a_i \in [0, 1]$. Informally, the marginal benefit of shirking must start out above the marginal cost at $a_i = 1$, and, as the level of compliance drops to zero, it must fall below the marginal cost: $B'(1) > C'(1)$, $B'(0) < C'(0)$. The point of intersection of the marginal benefit and marginal cost curves will identify the equilibrium level of compliance. The mathematical argument below characterizes the curvature conditions for shirking. I support the functional conditions with political motivations immediately following the proof.

We return to Eqn. 4 to find the value of $a_i$ that maximizes the value function. We are interested in equilibria where $V'$ is maximized for values of $a_i$ less than one but greater than zero. The first condition guarantees slippage, the second guarantees some minimal compliance. We make the following assumption: $V$ is quasi-concave in $a_i$.\footnote{People regularly assume that functions such as $V$ are quasi-concave. Here, it is not difficult to derive sufficient conditions for quasi-concavity, but the algebra becomes very messy. However, the conditions reduce to the following insights: (1) the change in marginal utility has to be large relative to the change in marginal probability of being caught or (2) the marginal probability of being caught must be small.} A stronger but sufficient condition, for the existence of an equilibrium, is that $V'$ is concave, which will be true if the second derivative of the
value function (Eqn. 3) is negative, which reduces to:
\[
[(1-\delta)+p\Delta]^2[U''((1-\delta)+p\Delta)-p''(U-X)\Delta]-2[(1-\delta)+p\Delta]p'[U'(1-\delta+p\Delta)-p'(U-X)\Delta] < 0.
\] (6)

**Proposition 4:** Given that the value function, \( V \), is quasi-concave, there exists a pure strategy equilibrium to this game.

**Proof:** Mas-Colell, Whinston, and Green, 1995, Proposition 8D3, p. 253. A pure strategy Nash equilibrium exists in the game \([I, \{S_i\}, \{V_i(\cdot)\}]\), where \( I \) is the set of players, \( S_i \) a set of strategies for each player, and \( V_i \) is a utility function for each player, provided \( \forall i \) (i) that the set of strategies \( S_i \) is a non-empty, convex, compact subset of some Euclidean space \( \mathbb{R}^m \), (ii) that \( V_i(\cdot) \) is continuous in all players' strategies, and (iii) that \( V_i(\cdot) \) is quasi-concave in \( s_i \). Condition (iii) holds given Eqn. 6. The game described meets all of these conditions. \( \square \)

By Proposition 4, we know that an equilibrium exists to the game in which players comply at some level \( a_i \geq 0 \). The following three propositions establish two sufficient conditions to guarantee that no full-compliance equilibrium exists: either the benefits of shirking must be large if everyone else is complying, or the likelihood of triggering the punishment regime must be small if one player shirks and the rest comply.\(^{24}\) The federation will exhibit shirking if either condition is present; players will “agree” upon some level of deviance, and no equilibria will include full compliance in the strategy set. While it is highly improbable that either the utility function or the probability of punishment is linear in the amount a player contributes (see the discussion below), Prop’s 5 & 6 each isolate the two sufficient conditions by restricting one of the two functions to linearity. We first consider how the utility function can contribute to shirking. We restrict our attention to symmetric equilibria, a condition we relax in Section 5.

**Proposition 5:** [Concave utility, linear probability of punishment] Suppose \( \frac{\partial p}{\partial a_i} = -\beta \). If (i) \( \frac{\partial^2 V}{\partial a_i^2} < 0 \), (ii) \( \frac{\partial V}{\partial a_i} \to 0 \) when \( \bar{a} \to 0 \) and (iii) \( \frac{\partial V}{\partial a_i} < -\beta \left( \frac{d(U(\bar{a}=1)-X)}{1-\delta} \right) \) at \( \bar{a} = 1 \), then the equilibrium

\(^{24}\)Note: Green & Porter (1984) prove a slippage result for concave utility functions. This model generalizes that result by considering a non-linear cost function.
will **not** support full compliance.

PROOF: See appendix.

The conditions in Proposition 5 are straightforward to interpret. The first condition guarantees that utility is maximized at some point. The second condition says that there are diminishing marginal returns from shirking; complying a little bit doesn’t hurt you all that much for very low levels of compliance, thereby ensuring some compliance in equilibrium. Under the third condition, if everyone else is complying fully, it becomes very tempting for you to shirk. If these conditions on the utility function hold, and the probability of punishment is linear in the level of compliance, then the equilibrium does not satisfy full compliance.

Proposition 6 inverts Proposition 5 to consider how the probability of punishment can contribute to shirking. Letting the utility function be linear in the level of compliance, Proposition 6 establishes a sufficient condition for shirking that depends upon the change in likelihood of triggering the punishment regime. The probability of punishment must approach zero as the participants reach near-full compliance, and for some level of compliance the probability of triggering the punishment regime for any marginal deviation must be high. The function itself must be non-linear, but may be convex or take an S-shape.

**Proposition 6:** [Linear utility, non-linear probability of punishment] Suppose \( \frac{\partial U}{\partial a} = -\alpha \). If an equilibrium exists, it does not support a strategy of full compliance if (i) \( \frac{\partial p}{\partial a} \to 0 \) as \( \bar{a} \to 1 \). Furthermore, if (ii) \( \exists \bar{a} \) for which \( \frac{\partial p}{\partial a} < \frac{-\alpha(1-\delta^{T+1})}{(\theta(\bar{a})-\gamma)|\delta-\delta^{T+1}|} \) at \( \bar{a} = \bar{a} \), then the equilibrium supports some minimal level of compliance.

PROOF: See appendix.

In Proposition 6, it is the first condition that guarantees slippage; if there is very little chance that one player will trigger the punishment regime if everyone else is complying fully, then the

---

25 I'm assuming that the utility function is concave, although it need not be to generate a quasi-concave value function.
agent will have little incentive to resist the urge to shirk. Furthermore, the proposition states that in order to guarantee some minimal compliance, there must be some point where an increase in one player’s level of shirking becomes very likely to trigger the punishment regime. If this condition is not present, the equilibrium strategy is full deviation and the federation probably would not exist.

Propositions 5 & 6 relaxed the linearity assumption on $U(a)$ and $p(a)$ in turn. Proposition 7 relaxes both assumptions, showing that weaker of the conditions in Prop’ns 5 & 6 is sufficient to guarantee non-existence of full contribution equilibria.

**Proposition 7:** [Concave utility, non-linear probability] If the probability of punishment, $p(a_i)$, is non-linear in $a_i$ and the utility function, $U$, is concave in $a_i$, then no full compliance equilibrium exists provided (i) $\frac{\partial U}{\partial a_i} < \frac{\partial p}{\partial a_i} \left( \frac{\delta [U(a_i)-1]-X)}{1-\delta} \right)$ at $\bar{a} = 1$. Furthermore, if (ii) $\exists \bar{a}$ s.t. $\frac{\partial p}{\partial a_i} < \frac{\partial p}{\partial a_i} \left( \frac{\delta [U(a_i)-1]-X)}{1-\delta} \right)$ at $\bar{a} = \bar{a}$, then the equilibrium will support some minimal level of compliance.

**PROOF:** Follows directly from Propositions 5 & 6.

In short, this section has shown that in public good contribution games, full compliance equilibria are a product of specific assumptions; generally, we cannot expect the participants to comply fully. If an equilibrium exists, it will allow for some slippage from full contribution, as long as the probability of punishment or the utility of deviation is non-linear.

The assumptions of concavity and non-linearity are more than mathematical conveniences; the discussion of the political environment on page 17 shows that they are reasonable ways to characterize interactions within a federal structure. Note that any one of Prop’ns 5, 6, or 7 can hold to make full compliance an unreasonable expectation in federal systems. With full information, the federation does not exhibit any structural defects; for all intents and purposes, it is as stable as any unitary state. With the addition of uncertainty, however, the decentralized structure of federalism—given that it exhibits at least one of the above conditions for non-linearity—is enough to generate competitive instability between regional governments, and between central and regional
levels. The cooperative equilibrium builds in a toleration for moderate deviation: even when
the union seems stable (when not in a punishment regime), all agents are allowed to deviate a
set amount. Encroachment, shirking, burden-shifting, and favoritism, in moderation, are to be
expected in all federal unions. From time to time, and due to random bad shocks (and not further
non-compliance on the part of the agents), the performance of the union falls below expectations;
with the primitive mechanism we have constructed, all will occasionally deviate fully, for some finite
period.

This result defines a role for more sophisticated institutions: they can constrain governments,
Madison-style, reducing the governments' ability to take advantage of one another. Institutions
could also correct the consequences of opportunism, perhaps through adjudicatory measures. In
short, institutions can reduce or eliminate the need for costly punishment regimes, making it
possible for the federal union to decrease the built-in tolerance for deviation and increasing its
stability.

5 Exit Options

The federalism context invites us to consider an extension missing from more generic analyses of
public goods games. In a federation, member governments have the option of exiting the union.
Assume that if a participant chooses to exit (the equivalent of not playing the game at all), its
expected utility is \( W \), a parameter that is exogenously determined and commonly known. In this
symmetric game, \( W_i = W_j \quad \forall \quad i, j \). The decision to exit is permanent. If a player exits, assume
the payoff to the remaining participants is \( \epsilon \geq 0 \), forever. The withdrawal payoff, \( W \), defines a
participation constraint. We can now discuss when that constraint will bind, and the effect of
having an exit option on the overall productivity of the union.

Let \( S \) denote the set of equilibrium strategies. Given \( s \in S \), we can define \( V_n(s) \) to be the
expected utility in normal play and \( V_p(s) \) as the expected utility at the start of the punishment
regime. Let \( s^* \) be the strategy that maximizes \( V_n(s) \) and \( \hat{s} \) be the strategy that maximizes \( V_p(s) \).
We focus finite period punishment strategies. We can write a strategy as \( s = (a, m, k) \) where \( a \) is
the contribution in normal play, \( m \) is the contribution during a punishment regime, and \( k \) is the
number of periods of punishment. It will be helpful, in the proofs of the next two propositions, to define the function $m(a, k)$ which equals the maximum contribution in each of $k$ punishment periods that would sustain a contribution level of $a$ in normal play. We will assume that $m(a, k)$ is continuous and differentiable where defined.\textsuperscript{26} We further assume that $m$ is defined in an open neighborhood around $a^*$, which is just a technical way of saying that you could support higher levels of cooperation in normal play, but it would not be efficient to do so. That is, slippage exists, as stated earlier.\textsuperscript{27}

**Proposition 8:** $V_n(s^*) > V_n(\hat{s})$: The contribution that maximizes expected utility at the start of the punishment regime generates does not maximize normal play expected utility.

**PROOF:** See appendix.

**Corollary 1:** If $V_n$ is concave, $\hat{a} < a^*$.

**PROOF:** See appendix, proof of Proposition 8.

**Proposition 9:** $V_p(s^*) < V_p(\hat{s})$: The strategy that maximizes normal play expected utility does not maximize expected utility at the start of the punishment regime.

**PROOF:** The proof follows the same logic as that for Proposition 8.

**Proposition 11:** We can fully characterize the equilibria as a function of $W$. (1) For $W < V_p(a^*)$, contribute $a^*$ in normal play. Expected utility is not affected. (2) For $V_p(a^*) < W < V_p(\hat{a})$, contribute a lar, where $\hat{a} < a_{bar} < a^*$, for a utility loss. (3) For $V_p(\hat{a}) < W < V_n(a^*)$ players do not participate in the union, for a utility loss. (4) For $V_n(a^*) < W$ players do not participate in

\textsuperscript{26} If no $k$-period punishment regime would sustain $a$, then $m$ is not defined.

\textsuperscript{27} I would like to thank an anonymous referee for pointing this out.
the union, for a utility gain. In particular, the introduction of an exit option translates into a net utility loss for moderate exit option utilities where \( V_p(a^*) < W < V_n(a^*) \).

PROOF: Table 1 and Figure 1 summarize the results.

<table>
<thead>
<tr>
<th>Case</th>
<th>Range</th>
<th>Contribution</th>
<th>E.U.</th>
<th>Utility Change</th>
<th>Union?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( W &lt; V_p(a^*) )</td>
<td>( a^* )</td>
<td>( V_n(a^*) )</td>
<td>none</td>
<td>yes</td>
</tr>
<tr>
<td>2</td>
<td>( V_p(a^*) &lt; W &lt; V_p(\tilde{a}) )</td>
<td>( \tilde{a} )</td>
<td>( V_n(\tilde{a}) )</td>
<td>decreases</td>
<td>yes</td>
</tr>
<tr>
<td>3</td>
<td>( V_p(\tilde{a}) &lt; W &lt; V_n(a^*) )</td>
<td>n/a</td>
<td>( W )</td>
<td>decreases</td>
<td>no</td>
</tr>
<tr>
<td>4</td>
<td>( V_n(a^*) &lt; W )</td>
<td>n/a</td>
<td>( W )</td>
<td>increases</td>
<td>no</td>
</tr>
</tbody>
</table>

1. \( W < V_p(a^*) \): The participation constraint does not bind. The union is not affected by the withdrawal option. Players have a strict preference to participate in the union. Any threat to use the withdrawal option to induce a higher payoff within the union is not credible. Furthermore, it cannot be sustained, as players could reduce the payoff from the punishment regime below the payoff from exit. For example, to induce compliance, participants may wage war on one another, a mechanism more costly—and therefore more effective—than simple non-compliance. Utility is unchanged.

2. \( V_p(a^*) < W < V_p(\tilde{a}) \): The constraint binds. Players choose the equilibrium contribution, \( \tilde{a} \) such that \( V_p(\tilde{a}) = W \). While the union can be sustained, it is not as productive as the union was without the exit option; the introduction of the exit option decreases equilibrium contribution levels. In the symmetric game, all players are strictly worse off by having the exit option.

3. \( V_n(\tilde{a}) < W < V_n(a^*) \): The constraint binds and makes the union impossible: participants hit the logical limit of how much they can increase \( V_p(\tilde{a}) \) by increasing the contribution during punishment periods. As in Case 2, the players' utility is strictly worse off by having the option to exit.

4. \( W > V_n(a^*) \): No union is possible, as the highest expected utility from participation in the union is less than the players' utility from exiting. Overall utility increases.

---

28The game-theoretic prediction is for no union. However, if we allow for cognitive constraints, this region may explain cases of unions that break apart at the first sign of trouble.
Corollary 2: If \( V_n(\hat{a}) > V_p(\hat{a}) \), then there is a discontinuous drop in utility at \( W = V_p(\hat{a}) \).

PROOF: See Figure 1. As exit option utility increases from \( V_p(a^*) \), agents can maintain the union by decreasing their normal play contribution \( a \) such that \( m(a, k) \) generates an expected utility at the start of the punishment regime that equals the exit option. At \( W = V_p(\hat{a}) \), players cannot increase their punishment play expected utility any more, and use their exit option. In so doing, their utility drops discontinuously from \( V_n(\hat{a}) \) to \( V_p(\hat{a}) \).

As long as the value of the exit option is greater than the expected utility at the start of the punishment regime, the addition of the exit option affects utility and often, the sustainability of the union. Without any option to quit the game, the optimal contribution in equilibrium, while not at full compliance, is as close as possible. Contribution is sustained by making the punishment regime severe enough that the players want to avoid it. Armed with an option to exit the union, players first compare \( W \) to \( V_n(a^*) \), the maximal expected utility of participation in normal periods. If the exit option exceeds this amount, naturally they will quit the union, as Case 4 describes. However, we see that even if the value of the exit option is less than the expected value of the union, the union might not be sustainable. As the exit option utility just surpasses the expected utility at the start of the punishment regime (Case 2), players might have some room to adjust the punishment regime utility upward by lessening the severity of the punishment. Players have many options for adjusting the expected utility at the start of the punishment period: they can reduce the severity of the single-period punishment (by complying more), they can reduce the duration of the punishment regime, or they can lower the threshold, thereby increasing the tolerance for non-compliance. While any one of these modifications raises the expected utility at the start of the punishment regime, they all will lower the maximal utility obtainable during normal periods. While these adjustments may rescue the union, they come at a cost: a less severe punishment induces less cooperative behavior from all members, reducing the utility from the union, and therefore reducing player utility. As

\(^{29}\)Only an increase in the level of patience could make the wait worthwhile, and in general, we assume that the discount parameter is exogenously determined.
Case 3 describes, players are limited in how much they can adjust the punishment utility. The presence of the exit option causes player utility to decrease and the union to dissolve.

Cases 2 and 3 are worrisome: players lose utility as a result of having a moderately-attractive exit option. At this point, institutional tinkering (beyond self-enforcing strategies) may help to sustain the union and prevent utility loss. For example, see Chen and Ordeshook (1994) for complementary analysis of constitutional secession clauses as a means to preserve federal unions.

6 Asymmetric Federal Arrangements

To this point the paper has only considered symmetric equilibria, where all agents are motivated in the same way by the federation and therefore all behave identically. Needless to say, federations are never perfectly symmetrical. This section shows how the general model can be modified to include power asymmetries in federations, considering several cases: (1) when one government has less to gain, overall, from being in the union than the other participants, (2) when a government gets less from its marginal contributions than others, (3) when a government feels the effect of the punishment regime more severely than other participating governments, and (4) when marginal deviations from one government are more likely to trigger the punishment regime, perhaps due to its size or visibility.\(^{30}\) To model asymmetric incentives, we now index the utility function and the utility from the punishment regime, \(U_i\) and \(X_i\), respectively, and allow the degree to which deviance affects the probability of punishment to vary from one agent to another: \(\frac{\partial u_i}{\partial a_i} \neq \frac{\partial u_i}{\partial a_j}\). The next corollary shows existence:

**Corollary 3:** Given quasi-concavity of \(V_i\) for all \(i\), there exists a pure strategy Nash equilibrium to this game.

**Proof:** Follows the same proof as Proposition 4.\(^{31}\)

\(^{30}\)Many more testable hypotheses can be generated if the general model is used as a foundation for assuming a specific functional form.

\(^{31}\)Again, concavity is a stronger but sufficient condition. The value function will be concave if:

\[
[(1 - \delta) + p\Delta]U'' \left( (1 - \delta) + p\Delta \right) - p\Delta(X) \Delta - 2[(1 - \delta) + p\Delta]U'(1 - \delta + p\Delta) - 2U'(U - X) \Delta < 0, \tag{7}
\]

for all \(i\).
The remaining four corollaries generate testable hypotheses about asymmetric federations. Corollary 4 considers the effect on the federal relationship if one government has a lower marginal utility from compliance and a lower overall utility from compliance.

**Corollary 4:** Suppose $\frac{\partial U_i}{\partial a_k} < \frac{\partial U_j}{\partial a_j}$ and $U_i(a) < U_j(a)$ $\forall a_i, \forall i \neq j$. Assume that $\frac{\partial p_i}{\partial a_k} = \frac{\partial p_j}{\partial a_j} = -\beta$. Then $a_i^* < a_j^*$.

PROOF: Assume $a^*$ is an equilibrium. Therefore, $\frac{\partial V_i}{\partial a_k} = \frac{\partial V_j}{\partial a_j} = 0$. By Eqn. 5 if $\frac{\partial V_i}{\partial a_j} = 0$ at $a_j^*$, then $\frac{\partial V_i}{\partial a_k}$, with its lower marginal utility and lower utility, could not be maximized at the same level of compliance. Since $V_i$ is concave, we see from Eqn. 5 that at $a_j^*$, $\frac{\partial V_i}{\partial a_i} < 0$. In equilibrium, the contribution by $i$ must be lower: $a_i^* < a_j^*$. □

This case demonstrates one form of independence; without resorting to a comparison of $W$, the participation constraint (the exit option), we see that some participants in the federation might get less out of participating than others. In this case, the federation must make allowances to tolerate a lower level of compliance for those members to maintain a stable federation.\(^{32}\)

**Corollary 5:** Suppose $U_i(a) < U_j(a)$ $\forall a, \forall i \neq j$. Then $a_i^* < a_j^*$.

PROOF: Follows from the proof for Corollary 4. Holding all else equal, the only term that varies in Eqn. 5 is the utility function. If $U_i(a) < U_j(a)$, then if $a_i^*$ is an equilibrium for $V_j$, then due to the value function’s concavity, at $a_j^*$, $\frac{\partial V_i}{\partial a_i} < 0$. Again, to restore equilibrium, the equilibrium contribution by $i$ must be lower: $a_i^* < a_j^*$. □

This result is similar to Corollary 4, however, here we see that even if the agents have the same marginal utility of compliance, if one government benefits less from the union than another, it will

\(^{32}\) Calvert (1993) considers a prisoner’s dilemma with random and private costs. While his model keeps the standard PD binary choice space for the agents (cooperate or defect), he shows that the efficient solution to the iterated PD is to allow agents with high costs to defect. He then shows how communication works as an institutional mechanism to keep the low-cost agents from mimicking the highs. In essence, the implication of his proposition and this article’s Corollary 4 (and 5) are the same: the less an agent stands to gain from participation in a joint project, the more allowances you must make for that agent to recoup some of the cost of participation by tolerating more deviance than you would allow from others.
be motivated to contribute less. Again, the federation must make allowances for this difference. We next consider a slightly different form of asymmetry, when the punishment regime harms some participants more than others.

**Corollary 6**: Suppose $U_i = U_j$ and $\frac{\partial p}{\partial a_i} = \frac{\partial p}{\partial a_j}$. If the utility gained from being in the punishment regime, $X$, differs between agents: $X_i < X_j \quad \forall i \neq j$, then $a_i^* > a_j^*$.

**PROOF**: Since $\frac{\partial U}{\partial a} < 0$ and $p(1) < p(0)$ and $p$ is monotonic in $a$, it can be shown from Eqn. 5 that as $X_i$ increases, $a_i^*$ decreases. Therefore, if $X_i < X_j$, then in equilibrium, $a_i^* > a_j^*$. \Box

Corollary 6 shows the deterrence of the punishment threat. As the utility from being in the punishment regime decreases, an agent will try to avoid the punishment regime by decreasing the probability that the system falls into the punishment mode. The only way a government can do this is to increase its own level of compliance, even though it suffers disutility from the surplus contribution.

Finally, we consider the effect on behavior of an asymmetric likelihood of triggering the punishment regime.

**Corollary 7**: Suppose $U_i = U_j$. If $\frac{\partial p}{\partial a_i} > \frac{\partial p}{\partial a_j}$, $\forall j \neq i$, then $a_i^* > a_j^*$.

**PROOF**: Again, consider Eqn. 5. Notice that $p_i(\bar{a}) = p_j(\bar{a})$; the probability of punishment is determined by the functional aggregation $F$ of all contributions, which affects all participants equally. If $\frac{\partial p}{\partial a_i} > \frac{\partial p}{\partial a_j}$ and $\frac{\partial V}{\partial a_j} = 0$ at $a_j^*$ $\Rightarrow$ $\frac{\partial V}{\partial a_i} > 0$. Therefore, $a_i^* > a_j^*$. \Box

In Corollary 7, one government is more likely to trigger the punishment regime than others, perhaps due to its visibility. Even though it receives no more utility from compliance than any other member government, it will comply at a higher level.

**ADD COROLLARY OF EXIT OPTION. Under construction.**

The original, symmetric model, although intentionally generic, would be dissatisfactory to any student of federalism—or of any public good contribution game, for that matter—who would reject
such a stark model when trying to describe a particular association. However, modelers should be warned against building in too many institutional or power assumptions, as one runs the risk of designing a model that fits only a single case, and perhaps limited to a particular point in time. It is preferable to have parameters that can be adjusted as needed, as this model allows. The move from a generic, symmetric model to one that captures power asymmetries forces us to think carefully about how to best model that asymmetry. For example, in most federations the center has coercive powers, be they as subtle as indirect influence over the constitutional court, or as apparent as a monopoly on military power. But how should we model coercive power? We must consider whether it imbalances the utility functions, and if so, does it make one government’s benefit from deviance higher or by reduce the overall potential benefit of being in the union? Or asymmetry could cause the relative harm felt during the punishment regime to increase to one player, or increase the cost of withdrawing from the union. Of course, a center’s size is not an unequivocal advantage: for example, any deviance on its part is more likely to affect the union, and therefore more likely to invite punishment, fitting with Corollary 7. As the corollaries above demonstrate, the model can be adapted to embrace many different power arrangements, and therefore be tailored to describe particular contribution games.

7 Discussion

Federal systems, at their heart constructed like a public good contribution game, will always exhibit some moderate creeping away from full compliance by its governmental participants; this opportunism can generate spillovers that can lead to squabbles between governmental units. Therefore, intergovernmental tension in federal systems is a product of the division of power between governments. It comes about because shirking on the federal bargain or constitution is initially beneficial to the governments, but the returns from shirking taper off; except with the primitive institution of the punishment regime, governments never deviate fully from the federal bargain, as long as their expected benefit from participation is greater than their outside opportunity. Also, as a government fails to comply, it becomes increasingly likely that the union will become destabilized, requiring the drastic measure of a punishment regime to reestablish it. Population heterogeneity may ex-
plain variances in the way some governments respond to the federal incentive structure, helping us to account for asymmetries, but any composition of citizens is likely to exhibit intergovernmental tension under the federal structure. Therefore, rather than try to eradicate all federal tension, to ensure the success of a federation we should work to manage the tension effectively.

The results in this paper raise a question of endogeneity of rivalry in newly established federations. The model predicts that political boundaries will become fault lines that ought to be monitored for the development of destabilizing intergovernmental tension. Prudent constitutional engineers will anticipate these lines of tension and design the incentive structure of the government accordingly. Specifically, the paper shows that we cannot always blame heterogeneity; nor is instability always the fault of immature political constitutions. Even in the United States, where we are largely content with our institutional structure, we still find intergovernmental rivalry. The role of institutions is not to eliminate all rivalry; the only way to do that is to have a unitary government, as this paper shows. Instead, institutions must manage the tension; rather than try to choke out all opportunism, the institution must ensure that the perceived value of the federation outweighs any outside option available to the governing participants. Therefore, the first challenge to the institutional structure is to guarantee the political feasibility of the federation.

To whatever extent possible, institutions can also help to increase the union’s efficiency. Federal efficiency has two meanings: First, it helps to determine the productivity of the union; in terms of the model, efficiency can determine the shape of the utility function, the way that compliance translates into return for each government. When evaluating the potential for stability in a federation, it is also important to consider a second form of efficiency: how costly is it for the union to line up incentives so that all members want to comply with the terms of the federal bargain? We might consider that the punishment regime is too extreme given that everyone knows that everyone else is complying (allowing for the tolerated slippage), and that any utility that falls below the threshold is due to stochastic shock, rather than deviance on the part of any player. But if the players would retreat from their commitment to the punishment regime, then the incentive structure would fail; compliance would shrivel and the union would become unsustainable.

Granted, the punishment regime is a costly mechanism for keeping all players honest. We
would prefer to have milder corrective mechanisms that avoid the sulky periodic retreat forecast by the punishment regime. Carefully designed institutions, generally some system of fragmentation and adjudication, fulfill the function of the punishment regime without the high cost; they set up incentives to ensure compliance. As long as they fulfill their functional requirements, they may be tailored to meet local traditions for greater chance of legitimacy. When the more sophisticated institutions are effective, they affect incentives in a manner far more efficiently than the punishment regime.33

In every federation, as in every public good game, exit options exist. At times they may be so undesirable, or so costly to pursue, that they hardly register as “options”. At other times, they are so fantastic that they foreclose any hope for union. But what effect does the exit option have when it is in the grey area, a mediocre to pretty good option? Section 5 extends our understanding of the influence of exit options by showing that if all players have a similar utility option, unless the option is so desireable that it is a “no-brainer” to leave, the presence of the exit option can only lower their utility.

Asymmetry manifests itself in a variety of ways, as corollaries 4 through 7 discuss. In designing institutions that acknowledge the inherent shirking in all federations, it is important to recognize the asymmetries that characterize the relationship between the participating governments. Knowing how to manage asymmetries effectively remains one of the greatest challenges to federal institutional design. Again, the problem reduces to flexibility: it seems unwise to permanently recognize particular asymmetries, as the utility functions of the agents may change over time. Instead, institutions must be accommodating as the era demands. Unfortunately, this flexible response requires a level of judgment, objectivity, and trust few institutions attain.

---

33 One of the most important innovations in institutional efficiency is monitoring of individual behavior. See Milgrom et al. (1990).
Appendix

Proof of Proposition 2: The infinite game full compliance equilibrium payoff and the payoff from deviations can be written as:

\[
C^\infty = U_i(1, 1-i) \left( \frac{1}{1-\delta} \right) \\
D^\infty = X \left( \frac{1}{1-\delta} \right)
\]

respectively.

In the finite punishment game, payoffs in the next \( T \) periods from cooperating and deviating are, respectively:

\[
C^T = U_i(1, 1-i) + \sum_{t=1}^{T} \delta^t U_i(1, 1-i) \\
D^T = U_i(0, 1-i) + \sum_{t=1}^{T} \delta^t X
\]

If \( C^\infty > D^\infty \), then \( \exists T^* \) s.t. \( C^{T^*} > D^{T^*} \). □

To derive comparative statics, we need to be more specific, and find the minimal \( T^* \) such that \( C^T \geq D^T \). For simplicity, let \( U(C) \) and \( U(D) \) represent the utility from cooperating and deviating, respectively.

\[
U_i(C) + \sum_{t=1}^{T} \delta^t U_i(C) \geq U_i(D) + \sum_{t=1}^{T} \delta^t X \\
\sum_{t=1}^{T} \delta^t U_i(C) - \sum_{t=1}^{T} \delta^t X \geq U_i(D) - U_i(C) \\
(U_i(C) - X) \sum_{t=1}^{T} \delta^t \geq U_i(D) - U_i(C) \\
\sum_{t=1}^{T} \delta^t \geq \frac{U_i(D) - U_i(C)}{U_i(C) - X} \\
\frac{\delta - \delta^{T+1}}{1-\delta} \geq \frac{U_i(D) - U_i(C)}{U_i(C) - X} \\
\delta - \delta^{T+1} \geq (1-\delta) \left[ \frac{U_i(D) - U_i(C)}{U_i(C) - X} \right] \\
\delta^{T+1} \leq \delta - (1-\delta) \left[ \frac{U_i(D) - U_i(C)}{U_i(C) - X} \right]
\]

(8)
For simplicity, let \( R = \delta - (1 - \delta) \left( \frac{U(D) - U(C)}{U(C) - X} \right) \). Then:

\[
\delta^{T+1} \leq R \\
(T + 1)ln\delta \leq lnR \\
Tln\delta + ln\delta \leq lnR \\
Tln\delta \leq lnR - ln\delta \tag{9}
\]

\( T^* \geq \frac{lnR - ln\delta}{ln\delta} \tag{10} \)

We now have a representation of the minimal number of punishment periods, \( T^* \), which we can use to generate comparative statics regarding \( U(C) \) and \( \delta \).

We use \( \frac{\partial R}{\partial U(C)} \) to find \( \frac{\partial T}{\partial U(C)} \). Using Eqn. 8:

\[
\frac{\partial R}{\partial U(C)} = \frac{[U(C) - X](1 - \delta) + (1 - \delta)[U(D) - U(C)]}{[U(C) - X]^2}
\]

Since we need only to sign the derivative, I suppress the denominator, as it is positive.

\[
\frac{\partial R}{\partial U(C)} = (1 - \delta)[U(C) - X + U(D) - U(C)] \\
= (1 - \delta)[U(D) - X]
\]

\[
\frac{\partial R}{\partial U(C)} > 0.
\]

Using Eqn. 8,

\[
\frac{\partial R}{\partial U(C)} > 0 \Rightarrow \frac{\partial T}{\partial U(C)} < 0. \tag{11}
\]

The number of punishment periods necessary to sustain cooperation varies inversely with the productivity of the union. \( \Box \)

We can also find the relationship between the patience of the players and the number of punishment periods. Note since \( R \geq \delta^{T+1}, \delta < R \Rightarrow T = 0 \). Therefore, in all interesting cases, where \( T > 0, R \leq \delta \). Coupled with the fact that \( \delta < 1 \), this implies

\[
\ln\theta < \ln\delta < 0. \tag{12}
\]
We are now ready to find $\frac{\partial T}{\partial \delta}$. Using Eqns 10 & 12, we can sign the derivative:

$$\frac{\partial T}{\partial \delta} = \frac{(\ln \delta)(-\frac{1}{\delta}) - (\ln \delta)(\ln R - \ln \delta)}{[\ln \delta]^2} \frac{\partial T}{\partial \delta} < 0.$$  \hspace{1cm} (13)

The minimal number of punishment periods varies inversely with the discount factor. □

**Proof of Proposition 3:** We restrict our attention to symmetric strategies. By using Eqns 4 and 5, we reconstruct $V'$ to read:

$$V' = \frac{\partial U}{\partial a_i} \left( \frac{1}{\delta} + p(\delta - \delta^{T+1}) \right) - \frac{\partial U}{\partial a_i} \left( (U - X)(\delta - \delta^{T+1}) \right) \frac{\partial V}{(1 - \delta) + p(\delta - \delta^{T+1})^2} (14)

If the probability of punishment and the utility from shirking are linear in the level of compliance, then we can write the derivatives of these functions as constants. Let $\frac{\partial U}{\partial a_i} = -\alpha$ and $\frac{\partial U}{\partial a_i} = -\beta$. We can now substitute these constants into Eqn. 14:

$$V' = \frac{(-\alpha)(1 - \delta + p(\delta - \delta^{T+1})) - (-\beta)(U - X)(\delta - \delta^{T+1})}{(1 - \delta + p(\delta - \delta^{T+1})^2) \frac{\partial V}{\partial a_i}} (15)$$

Recall that $p$ and $U$ are functions of $a_i$. We can show that

$$\frac{\partial V}{\partial a_i} |_{\delta=1} > \frac{\partial V}{\partial a_i} |_{\delta=0}$$

because $p(1) < p(0)$ and $U(1) > U(0)$, by the strategy profile and by assumption, respectively. Therefore, we have three possible cases: (1) If $\frac{\partial V}{\partial a_i} |_{\delta=1}$ is negative, then so is $\frac{\partial V}{\partial a_i} |_{\delta=0}$. The derivative is always negative, for all $a_i \in [0, 1]$. (2) If $\frac{\partial V}{\partial a_i} |_{\delta=0}$ is positive, then so is $\frac{\partial V}{\partial a_i} |_{\delta=1}$, and the derivative is always positive. (3) The derivative may start negative, $\frac{\partial V}{\partial a_i} |_{\delta=0} < 0$, and end positive, at $\frac{\partial V}{\partial a_i} |_{\delta=1} > 0$, and equal zero at some $a^* \in (0, 1)$. In case (1), no participant ever wants to comply more; agents are limited by the boundaries on $a_i$, so the equilibrium is $a^* = 0$. Case (2) represents the full compliance equilibrium; complying more is always better, so agents fully comply and $a^* = 1$. Case (3) allows for the possibility of an equilibrium with shirking. However, it can be shown\(^\text{34}\) that the

\[^{34}\text{In the linear case, the second derivative of the value function is written:} \]

$$V'' = V' - \frac{2p(\delta - \delta^{T+1})}{(1 + \delta) + p(\delta - \delta^{T+1})}$$

Therefore, if $V' = 0$, $V'' = 0$. 

35
second derivative is zero: $\frac{\partial^2 V}{\partial a_i^2} = 0$, and we have an unstable interior equilibrium. Note that while it doesn’t affect an individual’s utility to increase compliance marginally, it does raise all other players’ utility, because of the positive cross-partial: $\frac{\partial^2 V}{\partial a_i \partial a_j} > 0, \forall j \neq i$.

Therefore, the only stable equilibria are full compliance or no compliance. □

**Proof of Proposition 5:** Proposition 5 loosens the restrictions on the utility function from Proposition 3; we require only that the utility function be concave. Recall Eqn. 14:

$$V' = \frac{\frac{\partial U}{\partial a_i} ((1 - \delta) + p(\delta - \delta^{T+1})) - \frac{\partial p}{\partial a_i} ((U - X)\delta - \delta^{T+1}))}{((1 - \delta) + p(\delta - \delta^{T+1}))^2}.$$  

Proposition 5 is proven if (i) $\frac{\partial V}{\partial a_i} > 0$ at $\alpha = 0$ and (ii) $\frac{\partial V}{\partial a_i} < 0$ at $\alpha = 1$. Let $-\beta = \frac{\partial p}{\partial a_i}$ since the probability of punishment is linear. Substituting into Eqn. 14, the first condition is satisfied if:

$$\frac{\partial U}{\partial a_i} |_{\alpha = 0} \left( (1 - \delta) + p(0)(\delta - \delta^{T+1}) \right) + \beta \left( (U(0) - X)(\delta - \delta^{T+1}) \right) > 0$$

at $\alpha = 0$. If $\frac{\partial U}{\partial a_i} \rightarrow 0$ as $\alpha \rightarrow 0$, then it can be shown that the inequality holds. At $\alpha = 1$, the necessary condition becomes:

$$\frac{\partial U}{\partial a_i} |_{\alpha = 1} \left( (1 - \delta) + p(1)(\delta - \delta^{T+1}) \right) + \beta \left( (U(1) - X)(\delta - \delta^{T+1}) \right) < 0. \quad (16)$$

Since $p(1) > 0$ and is probably small, the inequality in Eqn. 16 holds if

$$\frac{\partial U}{\partial a_i} |_{\alpha = 1} < -\beta \left( \frac{(U(1) - X)(\delta - \delta^{T+1})}{(1 - \delta)} \right).$$

Therefore, at most $\frac{\partial U}{\partial a_i} |_{\alpha = 1}$ need be less than (and restoring $\frac{\partial p}{\partial a_i}$: $\frac{\partial p}{\partial a_i} \left( \frac{\delta(U(1) - X)}{1 - \delta} \right)$, □

**Proof of Proposition 6:** Here we allow the probability of punishment to be non-linear, while we keep the utility of deviation linear. Letting $\frac{\partial U}{\partial a_i} = -\alpha$, Eqn. 14 becomes:

$$V' = -\alpha \left( (1 - \delta) + p(\delta - \delta^{T+1}) \right) - \frac{\partial p}{\partial a_i} ((U - X)(\delta - \delta^{T+1})) \right) \quad (17)$$

As with Proposition 5, no full compliance equilibria exist if (i) $\frac{\partial V}{\partial a_i} > 0$ at $\alpha_i = 0$ and (ii) $\frac{\partial V}{\partial a_i} < 0$ at $\alpha_i = 1$. The second condition is satisfied if $\frac{\partial p}{\partial a_i} |_{\alpha_i = 1} \rightarrow 0$ as $\alpha \rightarrow 1$. However, it may also be true
that \( \frac{\partial p}{\partial a} \big|_{\delta=0} \simeq 0 \), in other words, the probability might not be convex, but instead might follow an S-shape. In this case, a sufficient condition to eliminate the possibility of a full compliance equilibrium is if there exists some \( \tilde{a} \in (0,1) \) for which the slope is sufficiently steep. Formally (dropping the denominator from Eqn. 17),

\[
0 < -\alpha \left( (1 - \delta) + p(\tilde{a})(\delta - \delta^{T+1}) \right) - \frac{\partial p}{\partial a} \left( (U\tilde{a} - X)(\delta - \delta^{T+1}) \right)
\]

\[
\frac{\partial p}{\partial a_i} \big|_{\tilde{a}} < -\frac{\alpha \left( (1 - \delta) + p(\tilde{a})(\delta - \delta^{T+1}) \right)}{(U(\tilde{a}) - X)(\delta - \delta^{T+1})}.
\]

While \( p(\tilde{a}) < 1 \), we can set it equal to 1 to make the r.h.s. as negative as possible. Then, it is sufficient for the derivative to be

\[
\frac{\partial p}{\partial a_i} < \frac{\left( \frac{\partial U}{\partial a_i} \right) (1 - \delta^{T+1})}{(U(\tilde{a}) - X)(\delta - \delta^{T+1})}. \quad \square
\]

**Proof of Proposition 8:** Let \( s^* = (a^*, m(a^*, k)) \). For any fixed \( k \), once a sustainable normal period contribution \( a \) is selected, it determines a unique \( m(a, k) \). It follows that \( m \) is strictly decreasing in \( a \). Therefore, with a fixed \( k \), we can define the payoff at the beginning of normal play solely as a function of \( a \): \( V_n^k(a) \). Since \( a^* \) maximizes \( V_n^k(a) \) and since \( m(a, k) \) is defined in the neighborhood of \( a^* \), it follows that \( V_n^{k'}(a^*) = 0 \).

Again keeping \( k \) fixed, we can also define \( V_p \) as a function of \( a \).

\[
V_p^k(a) = (1 + \delta \ldots + \delta^{k-1})U(m(a, k)) + \delta^k V_n^k(a)
\]

Taking the derivative with respect to \( a \),

\[
V_p^{k'}(a) = (1 + \delta \ldots + \delta^{k-1})U'(m(a, k))m'(a, k) + \delta^k V_n^{k'}(a)
\]

at \( a = a^* \), \( U' > 0 \), \( m' < 0 \), and \( V_n^{k'} = 0 \) which implies that the derivative is negative. Therefore, \( a^* \) cannot maximize \( V_p^k \). If \( V_n \) is concave, then \( \hat{a} < a^* \). \( \square \)

37
Bibliography


39


Figure 1: Expected Utility as a function of exit option utility. The shaded region represents utility loss from the exit option.