HARMONIZATION AND SIDE PAYMENTS IN POLITICAL COOPERATION*

Bård Harstad†

20 December 2004

Abstract

I study two regions that are negotiating an agreement to internalize externalities. Local preferences are local information, but reluctance is, in equilibrium, signaled by delay. Conditions are derived for when it is efficient to restrict the attention to policies that are uniform across regions - with and without side payments - and when it is efficient to prohibit side payments in the negotiations. While policy differentiation and side payments allow the policy to reflect local conditions, they create conflicts between the regions and thus, delay. If political centralization implies uniformity, as frequently assumed in the federalism literature, the results describe when centralization outperforms decentralized cooperation. But the results also provide a theoretical foundation for this uniformity assumption, and characterize when it is likely to hold.

Key words: Policy harmonization, side payments, bargaining, private information
JEL classification: H77, C78, D82

* I am grateful to Torsten Persson for carefully reading several drafts. I have also benefited from the help of Philippe Aghion, Alberto Alesina, David Austen-Smith, Patrick Bolton, Daniel Diermeier, Oliver Hart, Bob Staiger, Lucy White, and several seminar audiences. Thanks to Christina Lönnblad for editorial assistance, and Jan Wallander's and Tom Hedelius' Research Foundation and the Norwegian Research Council for financial support. The paper was earlier entitled "Uniform or Different Policies?".

† Kellogg School of Management, Northwestern University, 2001 Sheridan Rd, Evanston, IL 60208.

Email: harstad@northwestern.edu

1. Introduction

Consider the classical situation with regional public goods: Due to externalities, each of regions A and B contribute too little. The regions thus want to negotiate an agreement where they must both contribute more. The regions may be heterogeneous, but local preferences are local knowledge. In this simple context, I ask two simple questions. Can the regions benefit by constraining themselves to agreements with complete harmonization of policies? When will side payments improve the agreement?

International agreements and federal policies are often characterized by uniformity. The WTO applies specific formulas for tariff negotiations, and its "reciprocity principle" requires that concessions should be similar across countries. In the European Union, harmonization and uniformity are recognized as *the classic Community method.*[1] Article 94 TEC (Treaty of the European Community) calls for an *approximation of laws, regulations or administrative provisions*, which has led to an explosion of directives calling for uniform policies for e.g. the environment (McCormick, 2001). Since such uniformity cannot reflect local conditions, it is often criticized. Although Alesina, Angeloni and Etro (2005, p. 2) admit that harmonization is *a typical way in which the EU implements policies*, they argue that such "rigid" unions are inferior to "flexible" ones that allow for policy differentiation.[2]

The questions of differentiation and side payments are related as well as dependent. In fact, regional side payments may be interpreted as a form of policy differentiation, since they can be facilitated by negotiating federal taxes that differ (instead of being uniform) across regions. Side payments can also be explicit transfers between the regions, but they are most frequently facilitated by linkages between various issues combined in a package deal. The European Council, consisting of the head of states, typically negotiates

[1] De Burca and Scott (2000, p. 1).

[2] After a lot of similar criticism, the EU has indeed allowed more flexibility in recent years. Prior to the Maastricht Treaty, the above-mentioned Article 94 referred to *the approximation, or harmonisation.* Nugent (2003, p. 301) writes that *the word harmonisation was dropped to reflect the more flexible and less rigid approach that had developed towards differences in national standards and requirements.* De Burca and Scott (2000) argue that the EU has moved towards more flexibility as more members have been accepted.

such package deals. The Council of Ministers, however, has less discretion over alternative political issues, and side payments are difficult.[3] The allocation of tasks between the chambers may therefore determine whether side payments are possible. Also in other contexts, side payments are sometimes deliberately ruled out: Trade negotiations are seldom linked to the environment, for example. But because *linking issues together in 'package deals' can open the door to agreements by ensuring that there are prizes for everybody*, it is typically presumed that side payments can *sustain a vastly superior outcome compared to the agreement without side payments* and thus *side payments are needed to reach the best result*.[4]

Since both policy differentiation and side payments are presumed to be good, it is puzzling why they not always exist.[5] Before concluding on these questions, it is thus important to take a closer look at the trade-offs involved.

This paper presents a simple model where two regions negotiate on how to allocate the burden of producing a regional public good. The regions may be heterogeneous, but local information is local knowledge. The trade-offs related to differentiation and side payments are illustrated, and the best bargaining agenda is shown to depend on the amount of heterogeneity, the externality, and the value of the agreement.

Starting with the case where side payments are unavailable, the costs and benefits of differentiation are investigated. The regions have conflicting views over how the policy should be differentiated, since each region wants the other to contribute more. In equilibrium, the region with the highest value of the public good is most eager to settle the

[3]This is a well-known difference between the EU institutions. For example, Nugent (2003, p.357) writes that *The European Council has been instrumental in formulating some of the EU's grander compromises and linked deals* [while] *other EU institution...are ill-adapted to the linking of different policy areas* [because] *they certainly do not have the means.*

[4]The three quotas are borrowed from Nugent (2003, p. 357), Barrett (2001, p. 1847) and Cesar and de Zeeuw (1996, p. 158), respectively.

[5]For policy differentiation, Alesina, Angeloni and Etro (2005) conclude by admitting that *flexible unions may require complex organizational arrangements that may run into costs of complexity and difficulty of practical implementation.* Puzzled by the lack of side payments in international agreements, Cesar and de Zeeuw (1996) conjecture that *the reason might be that it is difficult...to determine the precise willingness to pay.*

agreement, and will have to contribute most. Such differentiation is typically efficient. But to obtain a better deal, each region would like to signal reluctance to participate in the agreement, and reluctance can be signaled by delay. Giving in early, a region reveals its impatience and high willingness to pay for the agreement, and it will have to contribute more. If the policies must be uniform, there is no desire to signal bargaining power by delay, since both regions will have to contribute by the same amount in any case. To evaluate whether policy differentiation is desirable, its value should be compared to the cost of strategic delay.

When side payments and differentiation are both possible, the situation is quite different. First, a region can "sell" its contribution against side payments. Thus, there are gains from trade. Second, the desired "direction of trade" depends on the region's value of the public good, so it can signal its type by its proposal. This make delay less necessary as a signaling device. For both reasons, political differentiation is always beneficial when side payments are possible. However, it might still be better to prohibit *both* side payments and differentiation, since this will eliminate all strategic delay.

By contrasting these two cases, the effect of side payments is isolated. As already noticed, side payments allow both gains from trade and a method for the regions to signal their types, implying that delay becomes less *necessary* as a signaling device. As a third effect, however, a reluctant region may force the other region to make side payments in addition to contributing more. With side payments, the conflict of interest is larger, and the *incentive* to signal reluctance by delay increases. When this effect dominates, side payments are detrimental to efficiency.

Besides providing normative recommendations for how to organize international negotiations, the paper contributes to several strands of literature. The literature on fiscal federalism, initiated by Oates (1972), typically compares decentralization and centralization under a quite restrictive assumption: if the policy is centralized, it *must* be uniform across regions. For international issues, this uniformity assumption is used to analyze the optimal and equilibrium size of nations (Alesina and Spoalore, 1997), the breakup of nations (Bolton and Roland, 1997) and political integration (Ellingsen, 1998). However, this assumption has recently been attacked by several scholars. Lockwood (2002) claims that the uniformity assumption is *not derived from any explicit model of government behavior*, and Besley and Coate (2003) argue that *since the reason for imposing the uniformity constraint seems unclear on both empirical and theoretical grounds, Oates' analysis is suspect.* This paper finds a rationale for uniform policies. But such harmonization requires a pre-commitment to uniformity, otherwise a reluctant region will demand favorable treatment. Since regions in a federal union are better able to make such commitments (by writing formal rules or using trigger-strategies in repeated interaction), more uniformity should be expected within federal unions, than across. Consequently, the paper provides a foundation for the uniformity assumption, and describes when it is likely to hold.[6]

The traditional "justification" for uniformity is that local preferences are private information. Alesina, Angeloni and Etro (2005, p.6) state that *information about countries' preferences is not publicly available.* Empirically, Mäler (1991) observes that *the control costs and environmental damage in one country are known to that country only.* But it is unclear why private information should imply uniformity. On the contrary, Bordignon, Manasse and Tabellini (2001) show that the policy is still differentiated in the optimal mechanism. But in a bargaining context, Alesina and Drazen (1991) predict a war of attrition between legislators trying to stabilize the economy, since the proposal-maker must bear the lion's share of the stabilization cost. As explained by Drazen and Grilli (1993), this is due to private information: accepting early stabilization reveals a politician's willingness to pay. Another politician will then require that the first bears most of the stabilization cost. Anticipating this, every politician is reluctant to propose stabilization, and stabilization is thereby delayed. Such haggling can, of course, arise in many contexts, and it may explain why it takes years to negotiate international agreements. The present paper adds to this literature by deriving conditions under which uniformity may help.

[6] A few other papers investigate how the political game may induce uniformity in alternative ways. Cremer and Palfrey (2000) show how a majority might vote for a federal mandate which might be too strict. Panagariya and Rodrik (1993) argue that a uniform tariff across industries might be optimal to reduce lobbying and tie the hands of politicians favoring certain groups. Using similar arguments, Besharov (2002) shows that uniform policies may be optimal to avoid costly lobbying. To my knowledge, this is the first paper investigating uniformity in a bargaining context.

There is a large literature on bargaining under private information (surveyed by Ausubel, Cramton and Deneckere, 2002). Because of the incentives to signal and screen, such negotiations are typically inefficient, and may involve delay. The model of this paper is similar to the endogeneous-timing models by Admati and Perry (1987) and Cramton (1992),[7] where each agent is allowed to delay as long as it desires before making its next offer. This paper contributes to the literature on bargaining under private information by investigating when efficiency may increase by imposing simple constraints such as uniform policies and a ban on side payments.[8]

The next section presents the model of the economy and the bargaining game. Section 3 describes the equilibrium when side payments are not possible, and investigates when a uniform policy is beneficial. Section 4 repeats this exercise for the case with side payments, while Section 5 compares the two cases and derives conditions under which side payments are good. The model is kept as simple as possible. While a bargaining game should only be interpreted as an example of how agents may negotiate, Section 6 finds the equilibrium to implement the most efficient "stable" and "robust" mechanism. Moreover, Section 7 argues that the results would hold under more general assumptions concerning the economic environment. Alternative interpretations of the results are discussed in Section 8. All proofs are relegated to the Appendix.

[7] Admati and Perry (1987) study a buyer-seller alternating offer bargaining game where the seller's type is known, while the buyer can be of two possible types. Cramton (1992) assumes there to be a continuum of types for each player.

[8] The paper also contributes to the literature on issue linkages, since side payments can be interpreted as such. This literature (e.g. Inderst, 2000) emphasizes the gains from trade and the distributive implications of adding another issue to the bargaining agenda. When there is private information, I find another rationale for issue linkages, as well as a counterargument. Since preferences can be signaled by the proposed combination of issues (or the "direction of trade"), using delay as a signaling device is less necessary. However, the second issue (side payments) may increase the conflict of interest and the incentive to signal bargaining power may increase. If the gain from trade is small, issue linkages reduce efficiency.

2. The Model

2.1. The Economic Environment

Consider a typical situation with regional public goods. Region A's marginal value of the public good is v_A, its contribution is measured by g_A, and the marginal cost of contributing is normalized to one. The externality e denotes the fraction of A's contribution that "crosses the border" to the benefit of region B. Since symmetric assumptions are made for region B, the level of public good in region A is $(1-e)\,g_A + eg_B$. In addition, let s be a (possibly negative) side payment from B to A. The regional utility functions are:

$$\begin{aligned} u_A &= v_A\left[(1-e)\,g_A + eg_B\right] - g_A + s \\ u_B &= v_B\left[(1-e)\,g_B + eg_A\right] - g_B - s \end{aligned} \tag{2.1}$$

To avoid trivial cases, assume that $1/(1-e) > v_i > 1$:[9] The first inequality ensures that no region will contribute without an agreement, the second that both regions would benefit from an agreement with equal contributions.[10] The positive externality[11] implies that the regions benefit from cooperation. Specifying (g_A, g_B, s) is equivalent to specifying (g, d, s), where $g \equiv g_A + g_B$ measures the *size* of the agreement and d measures how the policy is *differentiated* across the regions,

$$d \equiv g_B - g_A. \tag{2.2}$$

A *uniform policy* ($d = 0$) requires that both regions contribute the same amount, or equivalently, that the level of public good is the same in both regions.

Since the utilities are linear functions, the regions would like g to be infinitely large. If g must be financed by local taxes or loans, and the sum of the tax bases is normalized to 1, the budget constraint (when the regions can borrow from each other) is $g_A + g_B \leq 1$.

[9] In statements that may be true for either region, i denotes any of these, i.e. $i \in \{A, B\}$.

[10] I thus abstract from the problem of participation: If $v_i < 1$ were possible, i would not participate without differentiation or side payments (see e.g. Hoel, 1992).

[11] A negative externality can become positive by simply changing the sign of g_A and g_B: While the externality of pollution is negative, the externality of cleaning or *reducing* pollution is positive.

This constraint will bind in equilibrium, so that in any agreement,[12]

$$g_A + g_B = 1. \tag{2.3}$$

Besides reflecting the regions' joint budget constraint, a fixed size of the project is reasonable in several cases. In Alesina and Drazen (1991), legislators negotiate whether and how to stabilize debt, where the level of debt defines the required total contribution. Ratifying an environmental agreement may commit the EU to reduce its total pollution by a certain amount, but how the contributions should be allocated between countries remains to be negotiated. To fix ideas, suppose that A and B negotiate on whether and how to implement an agreement cleaning or reducing the total amount of emission by one unit. The amount of differentiation d measures how much B cleans relative to A. The externality e reflects the fraction of emissions crossing the border, and v_A and v_B measures A's and B's marginal value of clean air.

The constraint (2.3) allows us to focus on the distributive issues d and s. Combining (2.2) and (2.3), g_A and g_B become functions of d. Substituted in (2.1) gives

$$\begin{aligned} u_A &= (v_A - 1)/2 + d[1 - (1 - 2e)v_A]/2 + s \\ u_B &= (v_B - 1)/2 - d[1 - (1 - 2e)v_B]/2 - s. \end{aligned} \tag{2.4}$$

Total welfare can be defined as

$$u \equiv u_A + u_B = d(v_B - v_A)\left(\frac{1}{2} - e\right) + \frac{v_A + v_B}{2} - 1. \tag{2.5}$$

Equation (2.5) shows there exist potential benefits from differentiating the policy, whenever the regions are heterogeneous ($v_A \neq v_B$). If $e < 1/2$, pollution is mainly a local problem, and $d(v_B - v_A) > 0$ is optimal: the region with the highest value of clean air should reduce its emission most. If $e > 1/2$, most of the emission crosses the border and

[12] Without regional loans, an alternative assumption would be to let there be upper boundaries for *regional* contributions, $g_i \leq \overline{g}_i$. However, this would lead to an "unfair" bias towards uniformity, since the first-best would simply require $g_i = \overline{g}_i$, without capturing the intuition that the policy should be differentiated according to the v_is in optimum. Section 7 discusses how these assumptions could be relaxed.

$d(v_B - v_A) < 0$ is optimal: the region with the lowest value of clean air should reduce its emission most.[13]

However, A and B have conflicting interests in *how* the policy should be differentiated: Each region prefers that the other region contributes most. And, as motivated in the Introduction, local preferences are assumed to be local information. Region i only knows its own type $v_i \in \{\underline{v}, \overline{v}\}$, and the fact that the other region's type is either low ($\underline{v}$) or high ($\overline{v} > \underline{v}$) with equal probability.

This economy is formalized as simply as possible. However, the model can be extended in several ways: The utility functions may be concave and the constraint (2.3) relaxed, some heterogeneity may be observable, there might be heterogeneity in the costs of contributing, and the types may be arbitrarily distributed. Section 7 argues that while the analysis would then become more complex, the results would continue to hold.

2.2. The Bargaining Game

The bargaining game is quite standard. The regions make alternating offers over (d, s), A makes the first offer, time is continuous and the time horizon is infinite. Agreeing early is preferred to agreeing later, since i's present value of an agreement settled at time t is $\delta^t u_i$, where $\delta < 1$ is the regions' common discount factor. The *minimum* time between offers is negligible (small and approaching zero). I follow Admati and Perry (1987) and Cramton (1992) by relaxing the standard assumption that a region *must* make a proposal at a certain time. Timing is endogenous, as each region is allowed to delay as long as it wishes before making its next offer. This may be a reasonable assumption in international bargaining, where geographical distance may prevent efficient negotiations until the next meeting is scheduled.

Section 7 discusses how the bargaining game can be modified: Instead of allowing the regions to signal their types by delay, they may signal by proposing an inefficient (small) project. This would lead to similar results. However, Section 6 shows that the above

[13] Is it reasonable to allow $e > 1/2$? For the pollution example, this means that polluting plants are quite strategically located in each region. For international trade, it implies that a tariff reduction is mostly beneficial to foreigners.

bargaining game implements the most efficient "stable" and "robust" mechanism.

3. Without Side Payments: Uniform or Different Policies?

This section assumes that side payments are not possible ($s \equiv 0$). This may be reasonable if the negotiating ministers have little discretion over other political variables. Moreover, the results provide a benchmark case, which will later be contrasted to the case with side payments (Section 4) to evaluate the effect of the side payments themselves (Section 5). The first subsection describes the bargaining outcome when policy differentiation is possible, while the second subsection compares differentiation and harmonization.

3.1. The Outcome with Differentiation

This subsection states the bargaining outcome and explains how the negotiations will proceed.

Proposition 1: If the regions negotiate d, the unique sequential equilibrium satisfying the Intuitive Criterion has the following outcome (d, t):

B's type (3.1)

A's type	$\underline{v}$	$\overline{v}$
$\underline{v}$	$(0, t_2)$	(d', t_1)
$\overline{v}$	$(-d', t_1)$	$(0, 0)$

d', t_1 and t_2 are defined by:

$$d' = \frac{1}{2}\left[\frac{\overline{v}-1}{1-\overline{v}(1-2e)} - \frac{\underline{v}-1}{1-\underline{v}(1-2e)}\right] \quad (3.2)$$

$$\delta^{t_1} = 1 - \left[\frac{(1-\overline{v}(1-2e))d'}{\overline{v}-1+(1-\overline{v}(1-2e))d'}\right] \quad (3.3)$$

$$\delta^{t_2} = 1 - 2\left[\frac{(1-\overline{v}(1-2e))d'}{\overline{v}-1+(1-\overline{v}(1-2e))d'}\right] \quad (3.4)$$

The negotiations will proceed as follows. If region A is of high type, it will immediately propose equal contributions ($d = 0$). A high-type B immediately accepts. A low-type B, however, rejects A's offer and delays until time t_1 before suggesting (by proposing $-d'$) that A contributes most. This is immediately accepted by A. If A is instead of low type, it does not make any immediate offer. Instead, A delays until t_1 before suggesting (by proposing d') that B contributes most. A high-type B immediately accepts. A low-type B, however, rejects A's offer and delays until t_2 before suggesting equal contributions, which A immediately accepts.

In equilibrium, the region with the highest value of clean air will have to contribute most. If $(v_A, v_B) = (\underline{v}, \overline{v})$, for example, B is very eager to quickly settle the agreement. Since eagerness reduces B's bargaining power, A forces B to contribute most to the agreement ($d = d' > 0$). The amount of differentiation d' is determined by the difference in bargaining power. Since this agreement is quite attractive for A, A would like to pretend that it is of low type, even if this were false. To credibly signal its reluctance, a low-type A must delay in making an offer. The delay will be (exactly) so long that it would be too costly to afford for a high-type A.[14] A low-type B behaves in a similar way. Since only low types will delay, the agreement is settled earlier if more regions value it highly.

3.2. Uniform or Different Policies?

Suppose the two regions were committed to uniform policies ($d = 0$) should they ever reach an agreement. Then, the proceeding bargaining outcome would be simple. A would immediately suggest an agreement, and B would immediately accept, whatever are their types. No type would desire to imitate another type, so no delay would be necessary. Demonstrating bargaining power would be useless as the regions would have to contribute by the same amount in any case. With uniform policies and no side payments, the unique sequential equilibrium outcome is thus $(d, t) = (0, 0)$.

We can use the above results to characterize the expected and discounted total utility. In the case of uniform and differentiated policies, respectively, this can be written as:

$$u^0 = \frac{1}{4}(\overline{v} - 1) + \frac{1}{2}\left(\frac{\overline{v} + \underline{v}}{2} - 1\right) + \frac{1}{4}(\underline{v} - 1) \tag{3.5}$$

$$u^d = \frac{1}{4}(\overline{v} - 1) + \frac{1}{2}\left[\frac{\overline{v} + \underline{v}}{2} - 1 + \left(\frac{1}{2} - e\right)(\overline{v} - \underline{v})\, d'\right]\delta^{t_1} + \frac{1}{4}(\underline{v} - 1)\delta^{t_2} \tag{3.6}$$

[14] This is possible since the utility function $\delta^t u_i$ fulfills the single-crossing property.

By comparison, differentiation provides costs as well as benefits. The potential benefit is that the region with the highest value of clean air will reduce its emission most. The cost is that such an agreement is delayed. It is useful to define the expected value as

$$v \equiv \frac{\underline{v} + \overline{v}}{2},$$

and the heterogeneity by the relative difference in the two types' net value of a uniform agreement,

$$h \equiv \frac{\overline{v} - 1}{\underline{v} - 1} > 1.$$

Proposition 2: $u^0 \geq u^d$ if and only if condition (3.7) holds. This is more likely if the externality e is large, the heterogeneity h small, and the value v low.

$$h \left[2(v-1) \left(\frac{1-2e}{e} \right) - 1 \right] \leq 3 \tag{3.7}$$

The intuition is as follows. If the externality e is low, it is beneficial that the high-type region cleans most, since this will imply that the air is cleanest where this is most appreciated. Thus, the differentiation following from the bargaining game is valuable. If $e \approx 1/2$, however, it is of less importance where cleaning is located, since the amount of clean air will, in any case, be similar in both regions. The value of differentiation is then low. If $e > 1/2$, it would be optimal that the low-type region contributed most. In equilibrium, however, the high-type region ends up contributing most, since it has the lowest bargaining power. Requiring harmonization would then clearly be better. Thus, the benefit from a differentiated policy decreases in e. The cost turns out to increase. As e increases, each region benefits more from the other region's contribution, and the high-type becomes more tempted to imitate the low-type's strategy. To signal bargaining power credibly, delay must increase. In sum: if e increases, uniformity becomes better relative to differentiation.

Thus, it is necessary that $e < 1/2$ for differentiation to be good. If v then increases, there is an increase in the gains from cleaning domestically. The value of convincing the other region to contribute more instead decreases, and the high-type region becomes less tempted to signal bargaining power. Consequently, the delay decreases, and differentiation becomes better relative to uniformity.

If the heterogeneity h increases, the value of differentiation increases directly, which makes a differentiated policy better relative to a uniform one.

The chain of causation is admittedly somewhat more complex, however. If e, h, or v changes, so does the amount of equilibrium differentiation d'. And when d' changes, so do both the cost (delay) and the benefit (the amount) of differentiation. The proof of Proposition 2 shows that cost and benefit increase similarly when d' increases, and the two effects cancel.

In reality, d' may not be determined by negotiations alone. Economic or technological constraints may limit the extent to which the policy can be differentiated, such that

$$d \in [-D, D] \text{ for some } D \geq 0. \tag{3.8}$$

If this constraint is binding (i.e. if $d' > D$), it is easy to show that the outcome (3.1)-(3.4) continues to describe the equilibrium if only d' is replaced by D. And, as noticed in the previous paragraph, the amount of differentiation (d' or D) does not have an impact on whether a uniform policy is better. Proposition 2 holds in either case.

4. With Side Payments: Uniform or Different Policies?

Side payments may not be relevant if we consider coordination in a single issue. But they may be highly relevant when the coordinating countries are integrated also in other areas of policy, as in Europe today (Persson and Tabellini, 1995, p. 2000).

As issue linkages and logrolling become intrinsic in the political debate, some kind of side payments between the regions can be included and perhaps not excluded from the bargaining agenda. As already noticed, side payments can also be achieved by negotiating federal taxes that are different (instead of uniform) across regions. Thus, let the regions negotiate side payments as well as policy differentiation. The static Pareto frontier, drawn in Figure 4.1, shows that it is now necessary to recognize the limit (3.8) to differentiation. The bargaining game is similar to that above, but each proposal is now a pair (d, s).

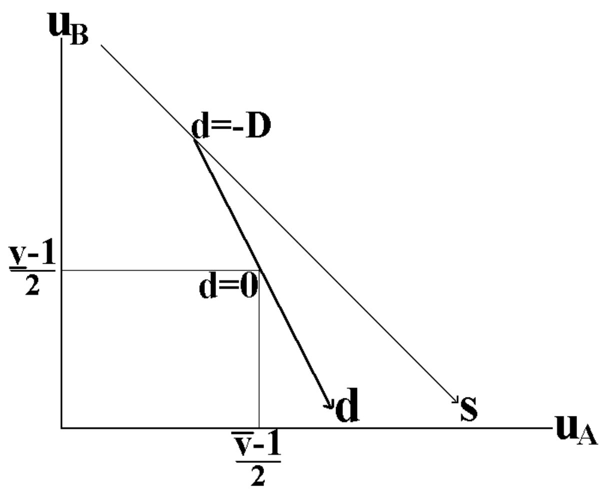

Figure 4.1: *The Pareto frontier when the agenda is* (d, s)*,* $e < 1/2$*,* A *is of high type and* B *is of low type.*

4.1. The Outcome with Side Payments and Differentiation

This subsection states the bargaining outcome and explains how the negotiations will proceed.

Proposition 3: If the regions negotiate d and s, the unique symmetric sequential equilibrium satisfying the Intuitive Criterion has the following outcome (d, s, t):

If $e \leq \frac{1}{2}$: B's type (4.1)

		$\underline{v}$	$\overline{v}$
A's type	$\underline{v}$	$(0, 0, t_2^s)$	$(D, \underline{s}, t_1^s)$
	$\overline{v}$	$(-D, \underline{s}, t_1^s)$	$(0, 0, 0)$

If $e \geq \frac{1}{2}$: B's type

		$\underline{v}$	$\overline{v}$
A's type	$\underline{v}$	$(0, 0, t_2^s)$	$(-D, \overline{s}, t_1^s)$
	$\overline{v}$	$(D, \overline{s}, t_1^s)$	$(0, 0, 0)$

t_1^s, t_2^s, $\underline{s}$ and $\overline{s}$ are defined by:

$$\delta^{t_1^s} = Min\left\{1 - \left[\frac{(\overline{v}-\underline{v})(1-|1-2e|\,D)}{2(\overline{v}-1)+(\overline{v}-\underline{v})(1-|1-2e|\,D)}\right], 1\right\} \quad (4.2)$$

$$\delta^{t_2^s} = Min\left\{1 - \left[\frac{2(\overline{v}-\underline{v})(1-|1-2e|\,D)}{2(\overline{v}-1)+(\overline{v}-\underline{v})(1-|1-2e|\,D)}\right], 1\right\} \quad (4.3)$$

$$\underline{s} \equiv \frac{1}{4}(\overline{v}-\underline{v}) - \frac{D}{4}[2-(\overline{v}+\underline{v})(1-2e)]$$

$$\overline{s} \equiv \frac{1}{4}(\overline{v}-\underline{v}) + \frac{D}{4}[2-(\overline{v}+\underline{v})(1-2e)]$$

The negotiations will proceed as follows. Suppose that $e \leq 1/2$. If region A is of high type, it proposes $(d, s) = (0, 0)$ at $t = 0$. A high-type B immediately accepts. A low-type B rejects A's offer and delays to t_1^s before it counteroffers $(-D, -\underline{s})$, which A accepts. If region A is of low type, it does not make any immediate offer. Instead, A delays to t_1^s before proposing $(D, \underline{s})$. A high-type B immediately accepts. A low-type B rejects A's offer and delays to t_2^s before it counteroffers $(0, 0)$, which A immediately accepts. If $e > 1/2$, the behavior is similar, but now a low-type region suggests that it will make the largest contribution itself. Naturally, the side payment to the low-type region must then be larger ($\overline{s} > \underline{s}$). In any case, the policy will be optimally differentiated in equilibrium.[15]

Just as in the previous section, a low-type region delays to credibly signal its bargaining power. A high-type region finds the low-type region's strategy unattractive, for two reasons. As before, a high-type region is less patient and cannot afford a delay. Second, if $e < 1/2$, the low-type region *pays* the other region to contribute most. A high-type region has a lower willingness to pay for such a "trade". Thus, a region signals its type by proposing a certain "direction of trade". If $D\,|1-2e|$ is large, the gains from such trade are large, the high-type is little tempted to imitate the low-type, and the necessary delay

[15] What determines the side payments? In the bargaining equilibrium, the side payments are set such that the utilities are equalized. If policies were uniform, side payments would go from the high-type to the low-type region, since the former benefits more from an agreement than the latter. But if one region is to contribute more than the other, it must be compensated. The net side payment will consist of the sum of these two forces. If $e < 1/2$, the high-type region should clean most. The two forces then pull in opposing directions, and it is unclear whether $\underline{s}$ is positive or negative. If $e > 1/2$, the low type should clean most. The side payment to the low type is then $\overline{s} > \underline{s}$, which is clearly positive.

to separate the two types is small. If $D\,|1-2e| \geq 1$, proposing a direction of trade is a sufficient signal: delay is not necessary and the bargaining outcome is first best.

4.2. When is Uniformity Better?

When both differentiation and side payments are possible, total expected utility can be written as

$$u^{ds} = \frac{1}{4}(\overline{v}-1) + \frac{1}{2}\left[\frac{\overline{v}+\underline{v}}{2} - 1 + (\overline{v}-\underline{v})\left|\frac{1}{2}-e\right| D\right]\delta^{t_1^s} + \frac{1}{4}(\underline{v}-1)\delta^{t_2^s}. \tag{4.4}$$

If uniform policies were required, but side payments were on the agenda, the outcome would be exactly as above by setting $D=0$. Define the resulting total expected utility as u^s. It is clear that u^{ds} increases in D, for two reasons. First, as D increases, it becomes possible to concentrate more of the cleaning to one region, and the gain from such differentiation increases. Second, it becomes more costly for the high-type region to imitate the low-type region's strategy, since this would imply inefficient differentiation. Thus, the need for delay is smaller. For these two reasons, it is always better to allow policy differentiation if side payments are on the agenda.

Proposition 4: $u^{ds} \geq u^s$ always.

This proposition does *not* imply that differentiation is good whenever side payments *can* be part of the agenda. It may be beneficial to prohibit *both* side payments and differentiation, since this ensures zero delay.

Proposition 5: $u^0 \geq u^{ds}$ if and only if both $D\,|1-2e| < 1$ and (4.5) hold. This is more likely if the heterogeneity h is small, the possibility to differentiate D small and $\left|\frac{1}{2}-e\right|$ small.

$$h\left(\frac{2}{1-|1-2e|\,D} - 3\right) \leq 3 \tag{4.5}$$

The basic intuition is as follows. If $e \approx 1/2$, it is of less importance where cleaning takes place, and there is little value in differentiation. If $e < 1/2$, most of the cleaning takes place in the high-type region in equilibrium, and the benefit of this is decreasing in

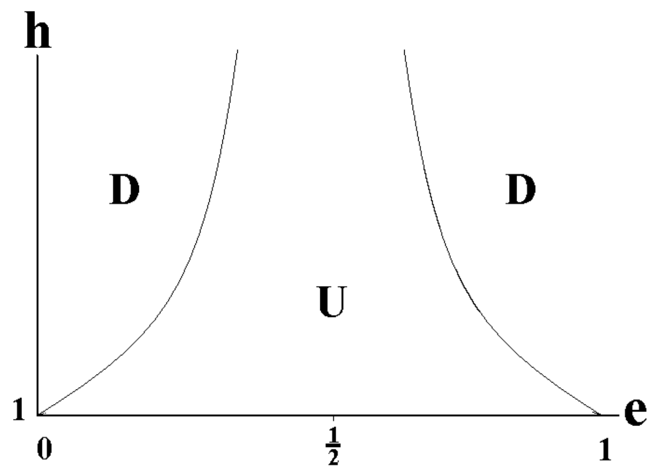

Figure 4.2: *In area U, it is optimal to prohibit both differentiation and side payments.*

e. If $e > 1/2$, optimal differentiation implies that the low-type region does most of the cleaning, and this benefit is increasing in e. In either case, the value of such differentiation is increasing in the heterogeneity h and the amount of differentiation, D. The potential cost of differentiation is delay, but this is decreasing in the gains from trade $D\,|1-2e|$, since such trade provides an efficient signaling device. If $D\,|1-2e| \geq 1$, there is no delay.

5. Are Side Payments Good?

-*Side payments are needed to reach the best result* (Cesar and de Zeeuw, 1996, p. 158).

Comparing Sections 3 and 4 reveals that side payments have three effects. First, there are "gains from trade", since one region can compensate the other for contributing more. For this reason, side payments are typically presumed to increase efficiency.[16] In addition, a region can signal its type by the proposed direction of trade, thereby making delay less necessary as a signaling device. But, as a third effect, a low-type region may force a high-type region to pay in side payments what it cannot pay in politics. Then,

[16]Cfr. the Introduction. From a game-theoretic perspective, however, it is not clear whether side payments are beneficial. Jackson and Wilkie (2003) show that the possibility to commit to side payments conditional on strategies may induce players to inefficiently tilt the equilibrium in their favor. Besley and Coate (2003) show that regions may be extensively expropriated if side payments are possible. In the bargaining context above, however, side payments would always be first-best if information were perfect.

bargaining power pays off more when side payments are available, and imitating the low-type's strategy becomes more tempting. There is then an increase in the *incentives* to signal bargaining power, and this effect might outweigh the reduced *necessity* to signal by delay.

Proposition 6: Suppose that $e \leq 1/2$. $u^d \geq u^{ds}$ if and only if (5.1) holds. This is more likely if the externality e is small, the heterogeneity h large, the value v of the agreement large, while the possibility to differentiate D small.

$$\frac{h-1}{h+1} \geq D\left(\frac{1-v(1-2e)}{v-1}\right) \tag{5.1}$$

The basic intuition is as follows. If D is low, the policy cannot be differentiated much and the gain from trading $(D-d')$ is small. Signaling a certain direction of trade is not a very convincing signal of type, and the delay is large. Since the policy cannot be differentiated a lot, bargaining power is not very useful without side payments. By allowing for side payments, however, the low-type region may force the high-type region to pay in side payments what it cannot pay in policy. The incentive to signal bargaining power increases, as does delay. Thus, if D is small, side payments are bad.

If the heterogeneity h is large, there is a great deal of differentiation d' even without side payments, and the gain from trading $D-d'$ is small. The difference in bargaining power is large, however, and the low type requires side payments from the high type. To signal bargaining power credibly, delay must increase when side payments are possible.

A large v and a small e make domestic contribution more valuable. Without side payments, regions are little tempted to signal bargaining power and there is little delay. Introducing side payments destroys the peace. Then, regions become more tempted to signal bargaining power in order to tilt the transfers in their direction, and delay may increase. Thus, side payments are good only if v is small and e is large. If $e > 1/2$, the policy is suboptimally differentiated without side payments. The gains from trade are then larger, and side payments are more likely to increase efficiency.

In short, unless the existing conflict between regions is sufficiently large, allowing side payments is detrimental to efficiency since it increases the conflict of interest and thus, delay. It follows that prohibiting side payments is always better if the policy must be uniform, i.e. $u^0 > u^s$.[17]

6. A Mechanism Design Justification

Since real-world negotiations are quite complex, any particular bargaining game and equilibrium should be considered as just an example of how the agents might negotiate. However, this subsection discusses two appealing properties of the above equilibria and shows that no procedure can achieve a better outcome with these properties.

First, notice that the final agreement d (for the case without side payments) or (d, s) (for the case with side payments) are identical to the bargaining outcome if information is complete (see the Appendix). The only difference is the delay. This feature is particularly important in the context of international agreements, where there is no third party enforcing the agreement. While the regions may reach an agreement today, it might not be binding tomorrow.[18] Any country may then declare the agreement invalid, after which each region contributes zero until a new agreement is formed. Since the regions' types are revealed at this stage, they will immediately negotiate a new agreement where d or (d, s) are as given above. Thus, only these agreements are "stable" towards such a unilateral request to renege the agreement.

Definition 1: An agreement is *stable* if no regions desire to renege the agreement ex post.[19]

Notwithstanding how the regions negotiate, their method can be replicated by a mechanism where honest revelation is an equilibrium. In our context, we can define a (direct) mechanism in the following way:

[17]In principle, the amount of side payments s could be limited by some constraint $s \in [-S, S]$. It is easily shown, however, that such a constraint would not change Proposition 6.

[18]Barrett (2001, p. 1836) writes that *the rules for international law allow countries to withdraw from an international treaty, at least after giving sufficient notice; and, as to reaffirm this freedom, nearly all treaties include an explicit provision for withdrawal.*

[19]This is different from traditional definitions of renegotiation-proofness (see e.g. Farrell and Maskin, 1989), precluding outcomes where *both* regions want to renegotiate.

Definition 2: Let $\widehat{v} = \{\underline{v}, \overline{v}\}$ be each player's strategy set. With differentiation but no side payments, a *mechanism* is a mapping $M^d : \widehat{v}^2 \to \mathbb{R}^2$, specifying an outcome (d, t) for each pair of possible types the regions may announce. With differentiation and side payments, a mechanism $M^{ds} : \widehat{v}^2 \to \mathbb{R}^3$ specifies an outcome (d, s, t) for each pair of announcement.[20]

Call a mechanism *stable* if the resulting agreement is stable (according to Definition 1). Mechanism design is often criticized because the optimal mechanism is typically sensitive to the agents' beliefs. Wilson (1985, p. 1101) claims that a desirable property of a mechanism is that it *does not rely on features of the agents' common knowledge, such as their probability assessments*. In the context of this paper, in particular, it might be hard to pin down exactly what one region knows about the other. Mechanisms robust to this criticism work for all priors the negotiators may possible have. In our context, this is equivalent to requiring that the mechanism is implementable in dominant strategies, and that honest revelation is an ex post equilibrium.

Definition 3: Call a mechanism *robust* if it is incentive compatible for all priors.

Proposition 7: The most efficient stable and robust mechanism M^d is implemented by the equilibrium in Proposition 1. The most efficient stable and robust mechanism M^{ds} is implemented by the equilibrium in Proposition 3.[21]

[20] There is no loss of generality by letting the mechanism be deterministic. Instead of letting the mechanism specify a probability p for the agreement to be formed, it can simply let the agreement be formed at date t, where $\delta^t = p$.

[21] How restrictive are the requirements that the mechanism must be stable and robust? If only relaxing the second requirement, the best mechanism dictates no delay if only one region announces low type, but a quite long delay if both regions do. The tradeoffs are qualitatively similar to those analyzed in this paper. Suppose that we instead relax the first requirement. Without side payments and assuming $e < 1/2$, the best separating mechanism implies that $d' = D$, while the delays are still given by (3.3)-(3.4). If $e \geq 1/2$, no mechanism can do better than a uniform policy. Proposition 2 remains unchanged. But with side payments, the first-best can be achieved just by adjusting the side payments.

7. Generalizations

In the above model, the utility functions are assumed to be linear and the constraint $g_A + g_B = 1$ is therefore imposed. This constraint "freezes" the size of the project while recognizing that the policy ought to be differentiated at the first best. Alternatively, the utility functions (or the v_is) could be concave functions of the amount of regional public good. The first-best would then be an interior solution and the constraint on g would not be necessary. Although the bargaining game would be too complicated to solve analytically, it is worthwhile to notice how the results would necessarily change. Crucially, different types would prefer different amounts of the public good. Even if uniformity were required, the regions would not have aligned preferences: The low type would prefer a smaller project (g) than the high type. The bargaining outcome would be a compromise. However, no type would like to imitate another type, so there would still be no delay with uniform policies. If policy differentiation were allowed, the low-type's proposal for a smaller project would not be tempting to imitate for the high type, so less delay would be necessary. The more the types differ, the larger would be the difference in their preferred size of the project, the less the high-type would like to imitate the low type, and the less delay would be necessary. Overall, the value of policy differentiation would increase in the heterogeneity, just as before. The other results above would continue to hold.[22]

It follows from the previous paragraph that a low-type region may signal its type by suggesting a less ambitious project. In fact, a low-type region can signal its type without delay by suggesting a sufficiently small project. To make the signal credible, the suggested project may be much smaller that what is efficient. This is also possible in the above model (but not an equilibrium).[23] Thus, the assumption that a region can

[22] If the v_is were concave, there would be a natural limit to how much the policy should be differentiated, even with side payments. The more concave were these functions, the more sensitive would the marginal value be to the amount of the public good, and the less the policy should be differentiated at the first best. To reflect this in a linear model, parameter D above should be smaller if the v_is are expected to be very concave.

[23] It is easily shown that, in the above model, a region would never signal its type by reducing g, since it is cheaper to signal by delay. The reason is that the other region can always suggest $g = 1$ in its next offer, while it cannot suggest to return to time $t = 0$ after some delay. Thus, the other region demands

delay as long as it wishes before making its next offer is not crucial. Without delay, the regions could propose a smaller project. In fact, if instead of delaying to time t, a region could permanently reduce the size of the project from 1 to δ^t, all the results above would continue to hold. Such an outcome would not be renegotiation proof, however, since the regions would both benefit from a more ambitious agreement.

The above bargaining game is simplified by assuming that $p_i = 1/2$, where p_i is the probability of region i being of low type, and assuming the types to be uncorrelated. Allowing for differences in the p_is and correlation, there would be no considerable change. However, for a sufficiently large p_i, there may be a pooling equilibrium since the other region would not find it worthwhile to screen region i. Then, the bargaining outcome might be uniformity, even if the regions are of different types. As shown in an earlier version of this paper, the set of parameters under which such pooling is an equilibrium is strictly smaller than that where uniform policies are optimal. A commitment to uniform policies (and no side payments) is still desirable for these parameters. The results of the paper thus survive, while the analysis will be more complicated.

All heterogeneity above is private information. However, it is straightforward to allow for some observed heterogeneity, such that $v_A \in \{\underline{v}_A, \overline{v}_A\}$ and $v_B \in \{\underline{v}_B, \overline{v}_B\}$. If the observed heterogeneity $(v_A - v_B)$ were large relative to the amount of private information $(\overline{v}_i - \underline{v}_i)$, differentiation would be more likely to be optimal. The trade-offs would be similar to those analyzed, however.

Finally, the heterogeneity may be related to the costs of contributing, instead of the values. If the value of the public good were v for both regions, while the cost of contributing were $c_i \in \{\underline{c}, \overline{c}\}$ instead of 1, the analysis would proceed in an almost identical way as compared to previously. Only one effect would disappear: As the externality increases above 1/2, it would still be desirable that the low-cost region contributes most. The value of differentiation would thus be independent of e. However, the cost of differentiation would still increase in e, since delay would increase in e for the same reason as before. Qualitatively, the above results would continue to hold.

less in the latter case.

8. Interpretations

The results can be interpreted in several ways. Obviously, they provide recommendations for how to organize international negotiations. The first subsection summarizes the results in this light. In addition, the second subsection argues that they explain why policies are typically more uniform within than across federal unions. Thus, the analysis provides a theoretical foundation for the uniformity assumption used in the literature. Based on this assumption, the final subsection explains that the results describe when centralization outperforms decentralized cooperation.

8.1. International Cooperation (A Summary)

Negotiating agreements between autonomous states or regions can be a difficult task. Two important questions are: To which extent should the policies be harmonized across regions? Should side payments be used?

Side payments? \ Policy differentiation?	no	yes
no	u^0	u^d
yes	u^s	u^{ds}

The answers turned out to be dependent. While policy differentiation is necessary to take heterogeneity into account, the regions will have conflicting views on how the policy should be differentiated, and this leads to inefficient bargaining (Proposition 1). By comparing the bargaining agendas in the first row of the above table, a uniform policy is better when the externality is large, the heterogeneity low, and the value of the project is low (Proposition 2). With side payments, there are gains from trade, but delay may prevail (Proposition 3). Comparing the two agendas in the second row of the above table, differentiation is always better (Proposition 4). However, it is better to prohibit both differentiation and side payments when the externality is large and heterogeneity low (Proposition 5). A vertical comparison between the two rows in the table reveals that it is efficient to prohibit side payments if the externality is low, the heterogeneity is large, and the value of the project is large (Proposition 6).

In short, while differentiation and side payments make the negotiations flexible and allow the policy to reflect local conditions, they create conflicts of interest and thus, delay. Moreover, the two are complements: With side payments, differentiation is always good; without differentiation, side payments are always bad.[24]

8.2. A Foundation for the Uniformity Assumption

While the analysis showed that harmonization may be the best policy, it is also clear that this requires the regions to commit in advance. Without commitment, a low-type region would easily benefit by proposing differentiation in its favor. One way of committing may be to use trigger strategies in frequent interaction, where regions stick to a limited bargaining agenda (without differentiation and/or side payments) to sustain future cooperation. Another way of committing is to write formal agreements, calling for harmonized policies for certain political issues. Either way, regions constituting a federal union should be better able to commit to uniform policies when this is the best solution. Hence, we should observe more uniform policies between regions forming a federal union than between regions that do not. This is similar to the uniformity assumption frequently made by the traditional literature on fiscal federalism, as discussed in the Introduction. The above analysis thus provides a theoretical foundation for this, and characterizes when the uniformity assumption is likely to hold. More importantly, the analysis provides both an explanation and a rationale for the empirical regularity that policies are more uniform within than across federations.

8.3. Decentralization vs. Centralization

As discussed in the Introduction, there is a large literature evaluating decentralization and centralization for both federal and international unions. This literature relies extensively

[24] Besides providing normative recommendations, the results can also be interpreted as positive predictions for how a rational EU would be organized. It is interesting to notice that while harmonization or approximation are recommended for the common market (Article 94 TEC) and the environment (Article 174), where the externalities are large, it is explicitly excluded for culture (Article 151) and public health politics (Article 152). This fit well with the theory. Moreover, as the EU expands, heterogeneity is likely to increase, and uniformity should be less desirable. Indeed, the EU has become more flexibility in recent years (De Burca and Scott, 2000, Nugent, 2003).

on two restrictive assumptions. First, as already discussed, the policy is assumed to be uniform whenever the instrument is centralized. Second, there is no coordination between regions if the policy is decentralized. As argued in the previous paragraph, this analysis provides some support for the first assumption. The second assumption seems more implausible, however. Even if the policy is decentralized, the regions certainly have incentives to cooperate whenever externalities exist. Suppose therefore that assumption 2 is relaxed. By decentralizing the policy, the regions are free to negotiate any agreement, including policy differentiation and maybe side payments. Due to private information, the negotiations are delayed, however. By centralizing the policy, in contrast, it becomes uniform across regions (according to the first assumption), and there is no delay. Thus, the comparison between centralization and decentralization is identical to that between uniform and different policies, analyzed in this paper.[25]

If side payments are not available, Proposition 2 describes when centralization is better than decentralization. Proposition 5 states when centralization is better in a context where the regions may use side payments. The analysis also suggests a case for partial decentralization: Comparing Propositions 2 and 6, a differentiated policy might be better than harmonization, but side payments may still be a bad idea. This will typically be the case if heterogeneity is large while the possibility to differentiate the policy is small. The best political regime is then to decentralize the relevant policy, while restricting the regions' discretion over side payments.

As compared to the traditional literature (e.g. Oates, 1972), certain results are confirmed: Centralization is better if heterogeneity is low and the externality large. The explanation is quite different, however. Proposition 2 also states that differentiation is better when the value of an agreement is large, since delay is then smaller. More important decisions should thus be decentralized, since the regions are then likely to negotiate more efficiently. This is similar to the result by Klibanoff and Morduch (1995). Other results are more at odds with the literature. In particular, it is the existence of asymmetric

[25] This trade-off is quite similar to that analyzed by Bolton and Farrell (1990), who study firms' entry on a market. While the cost of decentralizing this decision might be delay (as well as duplication when both firms enter), the benefit is that the most efficient firm is likely to enter first. A clumsy government, they assume, will immediately but randomly pick one firm.

information that makes the case for centralization in the model above. With complete information, the first-best is to differentiate the policy and thus allow decentralized coordination. With asymmetric information, instead, decentralized coordination is likely to be inefficient and centralization may be better. Moreover, the central government's uniform policies do not constitute a disadvantage, calling for more decentralization (as normally argued). Quite the opposite: It is the uniform policy which makes centralization potentially attractive, since it reduces the transaction costs of reaching an agreement.

Appendix

PROOF OF PROPOSITION 1: First, some notation is introduced. A *history* after N offers is the set of proposed and rejected offers: $H_N = \{d_N, t_N\}_N$. Let $\mathbf{H}_N$ denote the set of possible histories, define $H_0 \equiv (0,0)$, and let $\mathbf{H}$ be the set of all possible histories. A *pure strategy* for A is a rule f_A that says, whenever N is even, whether A should accept the previous offer or make a counteroffer d_{N+1} with delay $t_{N+1} - t_N \geq 0$, i.e. $f_A : \mathbf{H} \rightarrow \{accept, (\mathbb{R}, \mathbb{R}_+)\}$. Let A's *belief* $b_A : \mathbf{H} \rightarrow [0,1]$ denote the probability A puts on the state $v_B = \underline{v}$ after some history H_N. Similarly, f_B and b_B denote B's strategy and beliefs about A's type. At $t = 0$, $b_A = b_B = 1/2$.

Roughly speaking, a *sequential equilibrium* (Kreps and Wilson, 1982) is a set of strategies and beliefs such that after every history, each player's strategy is optimal, given its beliefs and the other player's strategy, and the beliefs are consistent with Bayes' rule. The *intuitive criterion* (Cho and Kreps, 1987) is a refinement restricting the beliefs outside the equilibrium. In essence, it requires that any action out of equilibrium that is beneficial for exactly one type, implies that beliefs place probability one on this type. In our context, this may be defined as follows:

Definition 4: Let $(d,t)_i$ denote the (expected) equilibrium outcome if i is of high type, given i's belief. Let $F_i \equiv \{(d,t) | (d,t) \succ_i (d,t)_i$ if and only if $v_i = \underline{v}\}$. The *intuitive criterion* requires that $b_j = 1$ after $i \neq j$ has taken some action leading to an outcome in F_i.

An earlier version of this paper let each region be of low type with probability p, and found that a pooling equilibrium might exist if $p \geq \overline{p}$ for some $\overline{p} > 1/2$. Since a pooling equilibrium is thus impossible when $p = 1/2$, I will restrict the attention to separating equilibria.

If information were complete, an argument similar to that of Rubinstein (1982) implies that the unique sequential equilibrium is the one defined by Proposition 1 with zero delay.[26] This level of differentiation is the best anyone can hope for in a separating equilibrium, since anything better would be rejected by the other type when the types are revealed.

Suppose that A is revealed to be of low type by making its last offer at $t_{\underline{A}}$. A high-type B will not be able to convince A that B is of low type. Thus, B accepts any $d \leq d'$, and will itself immediately propose d' if A's proposal is some $d > d'$. Should A understand that B is of low type, A would accept any $d \geq 0$. Thus, a low-type B maximizes its utility by proposing an offer in F_B which is acceptable to A if $b_A = 1$. That is, the offer must be unattractive to a high-type B and acceptable to a low-type A with beliefs $b_A = 1$:[27]

$$\underset{(d,t_2)}{Max} \ \frac{1}{2}\left[\underline{v} - 1 - (1 - \underline{v}(1-2e))\, d\right] \delta^{t_2} \ \text{ s.t.}$$

$$\begin{aligned} \frac{1}{2}\left[\overline{v} - 1 - (1 - \overline{v}(1-2e))\, d'\right] \delta^{t_{\underline{A}}} &\geq \frac{1}{2}\left[\overline{v} - 1 - (1 - \overline{v}(1-2e))\, d\right] \delta^{t_2} \\ d &\geq 0 \end{aligned}$$

[26] Note that an affine transformation of the utilities gives

$$\begin{aligned} \widetilde{u}_A &\equiv \frac{u_A}{[1 - v_A(1-2e)]/2} = w_A + d, \\ \widetilde{u}_B &\equiv \frac{u_B}{[1 - v_B(1-2e)]/2} = w_B - d, \text{ where} \\ w_i &\equiv \frac{v_i - 1}{1 - v_i(1-2e)} \end{aligned}$$

is region i's willingness to pay for the agreement in terms of d. In the Rubinstein (1982) alternating offer bargaining game, as the time between offers approaches zero, d will be set such that $\widetilde{u}_A$ and $\widetilde{u}_B$ are equalized:

$$d = (w_B - w_A)/2,$$

which gives the d in (3.1)-(3.2).

[27] Any other offer would either be less desirable for the low type, or also desirable for the high type. Thus, only this offer can constitute a separating equilibrium satisfying the Intuitive Criterion.

The solution is

$$\begin{aligned} d &= 0 \\ \delta^{t_2 - t_{\underline{A}}} &= \frac{\overline{v} - 1 - (1 - \overline{v}(1-2e))\, d'}{\overline{v} - 1}. \end{aligned}$$

Suppose instead that A is revealed to be of high type by making an offer at $t_{\overline{A}}$. A high-type B will not be able to convince A that B is of low type, and B accepts any $d \leq 0$, and will itself immediately propose $d = 0$ if A's proposal is some $d > 0$. A low-type B, on the other hand, maximizes its utility by proposing an acceptable offer in F_B. The problem is as before, and the solution is:

$$\begin{aligned} d &= -d' \\ \delta^{t_{\underline{B}} - t_{\overline{A}}} &= \frac{\overline{v} - 1}{\overline{v} - 1 + (1 - \overline{v}(1-2e))\, d'} \end{aligned}$$

Having found B's optimal strategy, let us turn to A. In a separating equilibrium, A makes an offer which only a high-type B accepts. If A is of high type, it cannot (by Definition 4) persuade B to believe that $b_B = 1$. Thus, a high-type A proposes $d = 0$ at $t_{\overline{A}} = 0$, which gives A the expected utility

$$\overline{u}_A = \frac{1}{4}(\overline{v} - 1) + \frac{1}{4}\left[\overline{v} - 1 - (1 - \overline{v}(1-2e))\, d'\right] \delta^{t_{\underline{B}}}.$$

The low-type A's problem is then to make an offer which is not attractive to a high-type A, but acceptable to a high-type B with beliefs $b_B = 1$:

$$\underset{(d, t_{\underline{A}})}{Max} \ \frac{1}{4}\left[\underline{v} - 1 + (1 - \underline{v}(1-2e))\, d\right] \delta^{t_{\underline{A}}} + \frac{1}{4}(\underline{v} - 1)\, \delta^{t_2 - t_{\underline{A}}} \delta^{t_{\underline{A}}} \ \text{ s.t.}$$

$$\begin{aligned} \overline{u}_A &\geq \frac{1}{4}\left[\overline{v} - 1 + (1 - \overline{v}(1-2e))\, d\right] \delta^{t_{\underline{A}}} + \frac{1}{4}(\overline{v} - 1)\, \delta^{t_2 - t_{\underline{A}}} \delta^{t_{\underline{A}}} \\ d &\leq d' \end{aligned}$$

The solution is:

$$\begin{aligned} d &= d' \\ \delta^{t_{\underline{A}}} &= \frac{\overline{v} - 1}{\overline{v} - 1 + (1 - \overline{v}(1-2e))\, d'} \end{aligned}$$

Combined, it follows that $\delta^{t_1} \equiv \delta^{t_{\underline{A}}} = \delta^{t_{\underline{B}}}$ and δ^{t_2} are such as defined in (3.3)-(3.4).

PROOF OF PROPOSITION 2: Define the *net* values of a uniform agreement as $\underline{n} \equiv \underline{v} - 1$ and $\overline{n} \equiv \overline{v} - 1$. Note that $\overline{n} = 2(v-1)h/(h+1)$ and $\left(1-\delta^{t_2}\right) = 2\left(1-\delta^{t_1}\right)$. $u^0 \geq u^d$ whenever the benefit from a differentiated policy is smaller that the cost of delay. Comparing (3.6) and (3.5),

$$\begin{aligned}
u^d &\leq u^0 \Leftrightarrow \\
\frac{1}{2}\left(\frac{1}{2}-e\right)(\overline{v}-\underline{v})d'\delta^{t_1} &\leq \frac{1}{2}\left(\frac{\underline{v}+\overline{v}}{2}-1\right)\left(1-\delta^{t_1}\right)+\frac{1}{4}(\underline{v}-1)\left(1-\delta^{t_2}\right) \Leftrightarrow \\
(1-2e)(\overline{v}-\underline{v})d' &\leq (\underline{v}+\overline{v}-2)\left(1-\delta^{t_1}\right)/\delta^{t_1}+(\underline{v}-1)2\left(1-\delta^{t_1}\right)/\delta^{t_1} \Leftrightarrow \\
(1-2e)(\overline{v}-\underline{v})d' &\leq (3\underline{v}+\overline{v}-4)\frac{[1-\overline{v}(1-2e)]d'}{\overline{v}-1} \Leftrightarrow \text{ } (d' \text{ cancels!}) \\
(1-2e)(\overline{n}-\underline{n})\overline{n} &\leq (3\underline{n}+\overline{n})[1-(\overline{n}+1)(1-2e)] \Leftrightarrow \\
(1-2e)(2\overline{n}+2\underline{n})\overline{n} &\leq (3\underline{n}+\overline{n})\,2e \Leftrightarrow \\
(1-2e)\,2(v-1)h &\leq (3+h)\,e \Leftrightarrow \\
h\left[2(1-2e)(v-1)-e\right] &\leq 3e \Leftrightarrow (3.7).
\end{aligned}$$

PROOF OF PROPOSITION 3: The proof is quite similar to the proof of Proposition 1, so only the key differences will be mentioned. The relevant concepts are defined in the pararell way. If information were complete, the unique sequential equilibrium would be given by Proposition 3 with no delay: This can be shown by a similar reasoning as that of Rubinstein (1982). As previously, no pooling equilibrium exists (when $p = 1/2$), so the attention can be restricted to separating equilibria.

Suppose that $e \leq 1/2$, and that A is revealed to be of low type by making an offer at $t^s_{\underline{A}}$. A high-type B will be unable to convince A that B is of low type, and will propose $d = D$ and $s = \underline{s}$, giving B utility $\overline{u}_B = [\overline{v}+\underline{v}-2+(\overline{v}-\underline{v})(1-2e)D]/4$. This offer equalizes and maximizes A's and B's utility of the agreement, and it is the best B can hope for. Thus, in considering A's offer, a high-type B accepts anything that would make B's utility at least as large as $\overline{u}_B$.

I will restrict the attention to symmetric outcomes (where $d = s = 0$ if $v_A = v_B$), which is satisfied if the agreement must be stable (defined in Section 6).[28] Then, a low-type B

[28] Why restrict the attention to symmetric/stable offers? If both regions are of low type, B can signal

proposes a stable agreement $(0, 0, t_2^s)$ acceptable to A but unattractive for a high-type B:

$$\underset{t_2^s \geq t_{\underline{A}}^s}{Max} \ \frac{1}{2}\left(\underline{v}-1\right)\delta^{t_2^s} \ \text{ s.t.}$$

$$\overline{u}_B \delta^{t_{\underline{A}}^s} \geq \frac{1}{2}\left(\overline{v}-1\right)\delta^{t_2^s}$$

The solution is:

$$\delta^{t_2^s - t_{\underline{A}}^s} = Min\left\{\frac{\overline{v}+\underline{v}-2+\left(\overline{v}-\underline{v}\right)\left(1-2e\right)D}{2\left(\overline{v}-1\right)} \ ,1\right\}$$

Suppose instead that A is revealed to be of high type by making an offer at $t_{\overline{A}}^s$. A high-type B will accept/propose $d = s = 0$, giving B utility $\left(\overline{v}-1\right)/2$. A low-type B prefers to propose the best agreement acceptable to A but unattractive for a high-type B. This is the stable offer $(-D, -\underline{s}, t_B^s)$ where:

$$\delta^{t_B^s - t_{\overline{A}}^s} = Min\left\{\frac{\overline{v}-1}{\overline{v}-1+\left(\overline{v}-\underline{v}\right)\left(1-(1-2e)D\right)/2}, 1\right\}$$

A makes an offer which only a high-type B accepts, namely $d = s = 0$ at $t_{\overline{A}}^s = 0$. This gives A the expected utility

$$\overline{u}_A = \frac{1}{4}\left(\overline{v}-1\right)+\frac{1}{4}\left[\overline{v}-1-\left(1-\overline{v}(1-2e)\right)D-\underline{s}\right]\delta^{t_B^s}.$$

The low-type A's problem is then to make an offer which is not attractive to a high-type A, but acceptable to a high-type B with beliefs $b_B = 1$. It can easily be shown that a low-type A will make a screening offer, i.e. $D, \underline{s}, t_{\underline{A}}^s$ where

$$\delta^{t_{\underline{A}}^s} = Min\left\{\frac{\overline{v}-1}{\overline{v}-1+\left(\overline{v}-\underline{v}\right)\left(1-(1-2e)D\right)/2}, 1\right\}$$

If $e > 1/2$, the proof proceeds in the same way, but since d changes signs in the optimal agreement, $(1-2e)D$ should be replaced by $\left|1-2e\right|D$. Combined, it follows that $\delta^{t_{\underline{A}}^s} = \delta^{t_B^s} \equiv \delta^{t_1^s}$ and $\delta^{t_2^s}$ are such as defined in Proposition 3.

this most cheaply by proposing that A contributes most ($d = -D$) and adjusting the side payments to equalize utilities. This would be quite expensive for a high-type B, so less delay is necessary. Allowing such an offer would make the outcome with differentiation more efficient, but the results would not change qualitatively. If small transaction costs were related to the side payments, A and B would prefer to renegotiate and set $d = s = 0$, when both are proven to be of low type. Hence, signaling by proposing $d = -D$ would not be credible, since the agreement would not be stable.

PROOF OF PROPOSITION 5: If $D\,|1-2e| \geq 1$, we know that the policy is optimally differentiated with no delay. Thus, assume that $D\,|1-2e| < 1$, and proceed as in the proof for Proposition 2:

$$\begin{aligned}
u^{ds} &\leq u^0 \Leftrightarrow \\
\frac{1}{2}\left((\overline{v}-\underline{v})\left|\frac{1}{2}-e\right| D\right)\delta^{t_1^s} &\leq \frac{1}{2}\left(\frac{\underline{v}+\overline{v}}{2}-1\right)(1-\delta^{t_1^s})+\frac{1}{4}(\underline{v}-1)(1-\delta^{t_2^s}) \Leftrightarrow \\
(\overline{v}-\underline{v})\,|1-2e|\,D &\leq [3\underline{v}+\overline{v}-4]\,\frac{(\overline{v}-\underline{v})\,(1-|1-2e|\,D)}{2\,(\overline{v}-1)} \Leftrightarrow \\
2\overline{n}\,|1-2e|\,D &\leq [3\underline{n}+\overline{n}]\,(1-|1-2e|\,D) \Leftrightarrow \\
\overline{n}(3\,|1-2e|\,D-1) &\leq 3\underline{n}\,(1-|1-2e|\,D) \Leftrightarrow \\
h\left(\frac{3\,|1-2e|\,D-1}{1-|1-2e|\,D}\right) &\leq 3 \Leftrightarrow (4.5).
\end{aligned}$$

PROOF OF PROPOSITION 6: Comparing (3.3)-(3.4) and (4.2)-(4.3), we notice that side payments reduce delay whenever

$$(\overline{v}-\underline{v})\,(1-|1-2e|\,D)\,/2 < [1-\overline{v}\,(1-2e)]\,d'. \tag{8.1}$$

Substituting for d', we observe that this condition always holds when $d' < D$. Then, side payments reduce delay (in addition to permit optimal differentiation) and are always good. But if $d' \notin [-D, D]$, d' must be substituted by D. Then, there are no gains from trade, and side payments are good if and only if (8.1) holds, which requires:

$$\begin{aligned}
(\overline{v}-\underline{v})\,(1-(1-2e)\,D) &\leq 2\,[1-\overline{v}\,(1-2e)]\,D \Leftrightarrow \\
(\overline{v}-\underline{v}) &\leq D\,[2-(\overline{v}+\underline{v})\,(1-2e)] \Leftrightarrow
\end{aligned}$$

$$h-1 \leq D\left[\frac{2e(h+1)}{v-1}-(h+1)\,(1-2e)\right] \Leftrightarrow (5.1).$$

This condition will always be satisfied when $d' < D$. If $e > 1/2$, the requirement for when side payments reduce delay is weaker and the gains from trade are larger. Thus, the larger is e, the more likely are side payments to increase efficiency.

PROOF OF PROPOSITION 7: Lets now calculate the most efficient mechanism that is stable and robust. The participation constraints are then fulfilled. This mechanism maximizes the total expected utility by minimizing delay, subject to these constraints and the regions' incentive constraints. Let t_0, t_1 and t_2 denote the time of the settlement when, respectively, neither, one and both regions announce low type. Since the game is symmetric, t_1 will no depend on which of the regions announces low type. With differentiated but no side payments, the problem is:

$$\underset{t_0,t_1,t_2\in[0,\infty)}{Max} u^d = \frac{1}{4}(\overline{v}-1)\delta^{t_0} + \frac{1}{4}(\underline{v}-1)\delta^{t_2} + \frac{1}{2}\left[\frac{1}{2}(\underline{v}+\overline{v}) - 1 + \left(\frac{1}{2}-e\right)(\overline{v}-\underline{v})d'\right]\delta^{t_1} \quad s.t.$$

$$\frac{1}{2}(\overline{v}-1)\delta^{t_0} \geq \frac{1}{2}[\overline{v}-1+(1-\overline{v}(1-2e))d']\delta^{t_1} \qquad (\overline{IC})$$

$$\frac{1}{2}[\overline{v}-1-(1-\overline{v}(1-2e))d']\delta^{t_1} \geq \frac{1}{2}(\overline{v}-1)\delta^{t_2} \qquad (\underline{IC})$$

$(\overline{IC})$ and $(\underline{IC})$ are the high type's incentive constraints when the other region announces high and low type, respectively. Both $(\overline{IC})$ and $(\underline{IC})$ must hold in an ex post equilibrium (which is equivalent to how I defined robustness in Section 6). It is easily checked that the low type's incentive constraints are not binding and can be ignored. The solution is that $t_0 = 0$, while t_1 and t_2 are set such that

$$\delta^{t_1} = \frac{\overline{v}-1}{\overline{v}-1+(1-\overline{v}(1-2e))d'}$$

$$\delta^{t_2} = \frac{\overline{v}-1-(1-\overline{v}(1-2e))d'}{\overline{v}-1+(1-\overline{v}(1-2e))d'} \Leftrightarrow (3.4).$$

With side payments and if $e \geq 1/2$, the problem is:

$$\underset{t_0^s,t_1^s,t_2^s\geq 0}{Max} u^{ds} = \frac{1}{4}(\overline{v}-1)\delta^{t_0^s} + \frac{1}{2}\left[\frac{\underline{v}+\overline{v}}{2} - 1 + (\overline{v}-\underline{v})\left|\frac{1}{2}-e\right|D\right]\delta^{t_1^s} + \frac{1}{4}(\underline{v}-1)\delta^{t_2^s} \text{ s.t.}$$

$$\frac{1}{2}(\overline{v}-1)\delta^{t_0^s} \geq \frac{1}{2}\left[\overline{v}-1+(1-\overline{v}(1-2e))D+\underline{s}\right]\delta^{t_1^s} \quad (\overline{IC})$$

$$\frac{1}{2}\left[\overline{v}-1-(1-\overline{v}(1-2e))D-\underline{s}\right]\delta^{t_1^s} \geq \frac{1}{2}(\overline{v}-1)\delta^{t_2^s} \quad (\underline{IC})$$

$(\overline{IC})$ and $(\underline{IC})$ are the high type's incentive constraints when the other region is of high and low type, respectively. Substituting for $\underline{s}$, it follows that $t_0^s = 0$, while t_1^s and t_2^s are set such that:

$$\delta^{t_1^s} = \frac{\overline{v}-1}{(\overline{v}-1)+\frac{1}{2}(\overline{v}-\underline{v})(1-(1-2e)D)} \text{ and}$$

$$\delta^{t_2^s} = \frac{(\overline{v}-1)-\frac{1}{2}(\overline{v}-\underline{v})(1-(1-2e)D)}{(\overline{v}-1)+\frac{1}{2}(\overline{v}-\underline{v})(1-(1-2e)D)} \text{ if } (1-2e)D<1.$$

$$t_0^s = t_1^s = t_2^s = 0 \text{ if } (1-2e)D \geq 1.$$

If $e > 1/2$, a similar maximization problem gives the same solution if only $(1 - 2e)$ is replaced by $|1 - 2e|$.

References

Admati, A. and M. Perry (1987): "Strategic Delay in Bargaining", *Review of Economic Studies* 54: 345-64.

Alesina, A., I. Angeloni and F. Etro (2005): "International Unions", *American Economic Review*, forthcoming.

Alesina, A. and A. Drazen (1991): "Why are Stabilizations Delayed?", *American Economic Review* 81 (5): 1170-88.

Alesina, A. and E. Spoalore (1997): "On the Number and Size of Nations", *Quarterly Journal of Economics* 112: 1027-56.

Ausubel, L., P. Cramton, and R. Deneckere (2002): "Bargaining with Incomplete Information", in R. J. Aumann and S. Hart (Eds.), *Handbook of Game Theory*, Vol. 3, Amsterdam: Elsevier Science B.V., Chapter 50: 1897-1945.

Barrett, S. (2001): "International Cooperation for Sale", *European Economic Review* 45: 1835-50.

Besharov, G. (2002): "Influence Costs in the Provision of Local Public Goods", Mimeo, Duke University.

Besley, T. and S. Coate (2003): "Centralized versus Decentralized Provision of Local Public Goods: A Political Economy Analysis", *Journal of Public Economics* 87 (12): 2611-37.

Bolton, P. and J. Farrell (1990): "Decentralization, Duplication, and Delay", *Journal of Political Economy* 98 (4): 803-26.

Bolton, P. and G. Roland (1997): ”The Break-up of Nations: A Political Economy Analysis”, *Quarterly Journal of Economics* 112 (4): 1057-90.

Bordignon, M., P. Manasse and G. Tabellini (2001): ”Optimal Regional Redistribution Under Asymmetric Information”, *American Economic Review* 91 (3): 709-23.

Cesar, H., and A. de Zeeuw (1996): "Issue Linkage in Global Environmental Problems", in A. Xepapadeas (Ed.), *Economic Policy for the Environment and Natural Resources: Techniques for the Management and Control of Pollution*, Cheltenham, E. Elgar.

Cho, I. and D. Kreps (1987): ”Signaling Games and Stable Equilibria”, *Quarterly Journal of Economics* 102: 179-221.

Cramton, P. (1992): ”Strategic Delay in Bargaining with Two-Sided Uncertainty”, *Review of Economic Studies* 59: 205-25.

Cremer, J. and T. Palfrey (2000): ”Federal Mandates by Popular Demand”, *Journal of Political Economy* 108 (5): 905-27.

De Burca, G. and J. Scott (2000): *Constitutional Change in the EU: From Uniformity to Flexibility?*, Hart Publishing Ltd.

Drazen, A. and V. Grilli (1993): "The Benefit of Crises for Economic Reforms", *American Economic Review* 83 (3): 598-607.

Ellingsen, T. (1998): "Externalities vs Internalities: A Model of Political Integration", *Journal of Public Economics* 68: 251-68.

Farrell, J. and E. Maskin (1989): "Renegotiation in Repeated Games", *Games and Economic Behavior* 1: 327-60.

Hoel, M. (1992): "International Environment Conventions: The Case of Uniform Reductions of Emissions", *Environmental and Resource Economics* 2: 141-59.

Inderst, R. (2000): "Multi-issue Bargaining with Endogenous Agenda", *Games and Economic Behavior* 30: 64-82.

Jackson, M. and S. Wilkie (2003): “Endogenous Games and Mechanisms: Side Payments Among Players”, *The Review of Economic Studies*, forthcoming.

Klibanoff, P. and J. Morduch (1995): "Decentralization, Externalities, and Efficiency", *The Review of Economic Studies* 62: 223-47.

Kreps, D. and R. Wilson (1982): ”Sequential Equilibria”, *Econometrica* 50 (4): 863-94.

Lockwood, B. (2002): ”Distributive Politics and the Benefits of Decentralisation”, *Review of Economic Studies* 69: 313-38.

McCormick, J. (2001): *Environmental Policy in the European Union*, Palgrave.

Mäler, K. G. (1991): "Environmental Issues in the New Europe", in Atkinson, A. B., and R. Brunetta (Eds.), *Economics for the New Europe*, New York University Press.

Nugent, N. (2003): *The Government and Politics of the European Union*, The MacMillan Press LTD.

Oates, W. (1972): *Fiscal Federalism*, Harcourt Brace: New York.

Panagariya, A. and D. Rodrik (1993): ”Political-Economy Arguments for a Uniform Tariff”, *International Economic Review* 34 (3): 685-703.

Persson, T. and G. Tabellini (1995): ”Double-Edged Incentives: Institutions and Policy Coordination”, *Handbook of International Economics*, Vol. III.

Rubinstein, A. (1982): "Perfect Equilibrium in a Bargaining Model", *Econometrica* 50 (1): 97-110.

Wilson, R. (1985): "Incentive Efficiency in Double Auctions", *Econometrica* 53 (5): 1101-15.