Abstract:

Why states would ever join federal unions when they could instead cooperate through IOs is puzzling. States form federations instead of international organizations when two conditions are present: the leaders that govern those states perceive a benefit from closer cooperation, and states that would have disproportionately valuable investments at stake in cooperation refuse to cooperate through simple alliances or customs unions and hold out instead for a federation. A case study of Australia’s union provides some empirical support.

The six states that today make up Australia unified under a federal constitution in 1901, just over 40 years on average after they became self-governing and effectively independent from Britain. Prior to federating, the states maintained separate armies and imposed high tariffs on each others’ goods. In the 1880s and ‘90s, their leaders tried several times to negotiate a customs union, following the desires of nearly all major interest groups in the region for a unified common market, but after several rounds of failed negotiations they concluded that it would be impossible for them to agree on a treaty. The following year the states opened negotiations on federation, and quickly agreed to a federal constitution. Why would states form federations when they could instead cooperate through international organizations such as customs unions or military alliances, or simply through ad hoc arrangements? I theorize that they form federations when they value cooperation but commitment problems make cooperation through international agreements unsustainable.

Beyond the historical – and perhaps contemporary – importance of federation for particular regions, understanding federation is important for the study of international organizations (IOs) and cooperation generally. I argue that there are some instances in which states are unable to cooperate without binding political institutions. Cooperation always requires states to divide among them the gains they reap by cooperating, but cooperation also changes the states’ bargaining power. As a consequence, a negotiated division that is a necessary part of a decision states make to begin cooperation may not be.

---

1 William Riker (1975) counts nearly half the human race living in a country that was formed by a federal merger among states that could otherwise have been independent. Examples include Argentina, Australia, Canada, Germany, India, Malaysia, Switzerland, the United Arab Emirates, and the United States of America, among others. Many more states, including nearly all Latin American countries, were once part of federal unions that failed, and most post-colonial African states were members of proposed federations. Processes that Alfred Stepan calls “coming-together” federalism (Stepan 1999, p. 21-22) may also be underway in Europe (Goldstein 2001) or in the world as a whole (Rodrik 2000; Wendt 2003).
appropriate to sustain cooperation once it has begun, and vice versa. Cooperation sometimes requires states to make investments in assets that are relationship-specific (more valuable while cooperation persists). States that are more heavily invested in relationship-specific assets than their partners find themselves increasingly, as cooperation persists, held hostage by the partnership, since the more they cooperate the more they have to lose if cooperation ends. These states’ bargaining positions erode over time, exposing them to extortion by their less-heavily-invested partners. This contracting problem can prevent states from cooperating even when cooperation is in their common interest. Federal constitutions are a way for states to contrive symmetry by making it so that ending cooperation is costly for all members. This resolves the commitment problem over how to distribute the joint benefits from working together, making all members confident that they can gain by cooperating.

My approach is to describe cooperation making as few assumptions as necessary in order to support a logical argument for why states in certain situations would choose a federation over cooperation via ad hoc arrangements or through an international organization. The theoretical argument is therefore like a possibility theorem (Stone 2002); my goal is to show a how a logically-consistent set of assumptions can completely explain an outcome where states reject international organization in favor of federation, something that has not yet been done in the study of federal unions.

I proceed in five parts. First, I describe the problem of how states reap and then divide among them the gains from cooperation. Second, I show a problem in how states commit to particular divisions. Third, I explain how federal institutions can solve the commitment problem. Fourth, I address the preferences states have and show how, when bargaining over institutions, they can settle on a federation instead of an international organization. Fifth, I conclude with an abbreviated case study of the Australian federation and some general observations about international organizations today.

Reaping and dividing the gains from cooperation

Although there is a rich literature examining the workings and stability of federations (Weingast 1997; Filippov, Ordeshook and Shvetsova 2003), and a growing literature that examines the construction of federal institutions over time once leaders have already chosen to federate (Goldstein 2001), few scholars have directly addressed the issue of when and why leaders choose to take their states into federations in the first place. William Riker (1975) and Alberto Alesina and Enrico Spolaore (1997; 2003) theorize that states in a federation will be more secure or wealthier than states that remain self-sufficient. These gains from cooperation can arise for a variety of reasons. Alesina and Spolaore begin with the premise that states produce some public good with a fixed cost and increasing returns to scale. States therefore benefit from size, since if they are larger they can spread the cost of the good across more taxpayers. This logic can drive federal unions among states as well, since leaders may seek to capture the cost savings that come with size (Alesina, Angeloni and Etro 2001).

One possible public good that states can provide more cheaply by being bigger is security. National defense is often described as nearly a pure public good (Olson 1965), and security threats may provide an incentive for states to merge (Alesina and Spolaore 2003, p. 111; Parent 2006). Security interests form the basis of William Riker’s causal
theory of federations, as he argues that states form federations when they have a common interest in either common defense or common offense (1964; 1975; 1987). Implicit here is the premise that states can at times make themselves more secure (or make themselves as secure for cheaper) working together than working separately. This might be because states can specialize in particular military tasks to achieve economies of scale (Lake 2001), or simply that states can combine their efforts to overpower a common rival.

Another possible public good is a large common market. The more people are enclosed within a single market, the more opportunities they will have for gains from specialization and exchange (Smith 1776). In this view a large market is like a public good in that each additional citizen added to the market actually increases the value of the market to every other citizen, on average (Alesina and Spolaore 2003, p. 4).²

Some empirical studies, however, have failed to find clear evidence that size matters, since, for example, small countries seem no worse off on average than large countries, by a variety of measures (Rose 2006). This may be because there are many ways for states to enjoy the benefits of size without actually becoming bigger, or joining federations with their neighbors. States can achieve gains from scale simply by cooperating with other states, either through informal cooperation or international organizations. Customs unions, military alliances, and other functional institutions abound; states do not face a binary choice between, on the one hand, federal integration, and, on the other hand, isolation and self-sufficiency. A complete theory of federation, therefore, cannot logically rest only on the observation, however true, that the size of a region for which a public good is produced matters.

States can produce more security or wealth, or both, working together than they can separately, but gains from joint production can be divided in any of a number of ways. Even if the natural technology of joint production creates some particular allocation of gains, such as when market forces distribute wealth among members of a common market, states can find other ways to redistribute. Most goods are themselves divisible, so that, for example, states can allocate among themselves the costs of military preparations for common defense, but even when they are not they can be linked to other issues through side-payments.³

When deciding how to divide the gains among them, states have a range of possible decisions they can make that are consistent with preserving their relationship. The boundaries of this range are determined by how well each member would fare if the cooperation were to come to an end, so that states will continue to cooperate as long as they are each, individually, at least as well-off by staying as by walking away. As a result, the actual division of gains – the policy decisions that the states make – depends

² Friedman (1977) argues that competing political authorities along a single trade route will overtax the route, while competing political authorities along parallel trade routes will undertax those routes; mergers allow rulers to extract the optimal, revenue-maximizing tax. Bolton and Roland (1997) assume that states merge for (unspecified) efficiency gains and analyze the consequences of regional income equality and redistribution. Larger size may also provide states with insurance against shocks to particular regions (Oates 1972).

³ For a concrete example of negotiations that make extensive use of side-payments in response to states’ outside options, see Mayer (1998) or Gruber (2000) on the NAFTA.
on the outside option that each state has. This is a common way of thinking about bargaining, as for example, in pioneering work by Howard Raiffa (1982) that describes how bargains are shaped by each side’s best alternative to a negotiated settlement. Where exactly the states decide to settle within the range of options acceptable to all depends on their relative skills in bargaining and negotiation.

States that would be relatively well-off without cooperation, those states that would be militarily secure and have access to large markets even without cooperating with partners, have a better outside option than states that would fare poorly without cooperation. States with better outside options need a more favorable division of the benefits in order to entice them to cooperate; they are in a better bargaining position. States, furthermore, have the ability to adjust the distribution of benefits among them in order to rebalance the division of gains as their outside options change. If a state’s outside option improves over time, it will be in a position to extract a better bargain from its partners. If, on the other hand, its outside options worsen, it will find itself getting a poorer and poorer division of benefits over time.4

Figure 1: Distributing joint gains.

![Figure 1: Distributing joint gains.](image)

4 The general recognition that outside options determine a state’s bargaining power often leads states to take steps to manipulate their outside options. Albert Hirschman referred to outside options as the power actors get from their ability to exit agreements (1970), and noted that Germany in the 1930s was able to manipulate other states’ outside options in ways that made it costly for central and southern European countries to leave the German economic sphere of influence (Hirschman 1945). Similarly, Lloyd Gruber (2000) argues that the value of outside options can be subject to manipulation when strong states find ways to worsen their weaker partners’ alternative to cooperation. A state can get a better deal for itself in cooperation if it has partners who would not do as well for themselves outside of a deal.
Figure 1 illustrates the logic of how states bargain over the gains from cooperation. The horizontal and vertical axes represent the payoffs to two different states. If they choose to work together, together they move from the point labeled “payoff to no cooperation” to the point labeled “payoff to cooperation.” Cooperation makes both better off: the one state can get more by cooperating than it could no cooperating (it moves to a payoff that is higher along the vertical axis) and the other state also gets more by cooperating (it moves to a payoff that is further to the right along the horizontal axis). Because the two states can redistribute the gains from cooperation among themselves, they can achieve any payoff along the 45 degree line going through the “payoff to cooperation” point. Assume that the transfer is negotiated at the same time that the states decide whether or not to work together, so that the two states simultaneously decide whether to cooperate or not and, if they do cooperate, how they will divide the gains.

The outcome they choose will be somewhere along the diagonal line segment in Figure 1. Neither state will accept a bargain and transfer that leaves it worse off than the payoff to no cooperation, so the line segment does not extend beyond the payoffs to the two states of no cooperation (the two dashed lines). Each state, of course, would ideally prefer a solution in which they work together and it gets all the benefits to itself – so that the one state prefers a solution near the upper-left of the 45 degree line while the other state prefers a solution in the lower right.

When the states bargain over what cooperation will look like – that is, what point along the diagonal line they will choose – the bargaining range is constrained by the minimal outcome each side will accept, which is defined by what each side would get if there were to be no agreement. The argument I will make is about how adjustment costs influence the bargaining range and sometimes make it disappear entirely, meaning no agreement is possible. For this reason, it is unnecessary to model the bargaining process itself. Assuming simply that the states always agree to split the difference between the two minimally-acceptable outcomes,^5^ the states end up at the point labeled bargaining outcome.

Committing to divisions

When a state cooperates with partners, it builds up assets that lose value if cooperation ends. If the act of ending cooperation is itself costly because it erodes the value of relationship-specific assets, then the state’s outside option is worse than simply the static value of self-sufficiency. These assets can be military or economic. A state that builds a navy but relies on allies for ground forces, or that builds an army but relies on allies for the logistical ability to transport it to battlefields, has specialized. If it loses

---

^5^ This assumption that states split the different – which here is equivalent to the Nash Bargaining Solution – is a common way to represent the outcome of a bargain where the bargaining process itself is unimportant to the theory is a useful way to examine contract negotiation (Watson 2002). The Nash Bargaining Solution is the division the states would have ended with in a number of non-cooperative bargaining games, such as an Osborne-Rubinstein bargaining game in which discount rates approached zero (Osborne and Rubinstein 1990).
the ally, it must pay the costs to make up the functions the ally had previously served (Lake 1999). Established procedures, joint training, and informal contacts between militaries are also relationship-specific assets that raise the costs of losing a partner (Weber 1997; Wallander 2000).

Security investments that depend on territory controlled by an ally are relationship-specific as well. During the early Cold War, American land-based, intermediate range ballistic missiles could not reach targets in the Soviet Union from the continental United States, and so were only valuable as a deterrent as long as they could be stationed on the territory of allies who were closer to the Soviets. Similarly, Soviet short and medium range missiles were most useful as a deterrent if they could be stationed on Cuba. The loss, to either side, of the ability to use allied territory would have made the missiles far less useful.

Economic forces can also create relationship-specific investments. Trade patterns in a common market can make a state dependent on its trading partners in a way that would force the state to make a costly readjustment if its trading partners were to sever ties. Such a readjustment may be more costly than the gains from trade in the first place (Alt, Carlsen, Heum and Johansen 1999). Consider a small state that specializes in its comparative advantage with respect to a trading partner. If that trading partner were to cut off economic relations, the state may have to entirely restructure its economy around trading with other partners or around producing for the home market in autarky (McLaren 1997).

Some economic assets related to trade (in addition to the production that trade itself creates) are also relationship-specific. Jeffrey Frieden (1994) notes that mines are site-specific, and so an ownership stake in a mine in a foreign country only generates revenue if the mine is not expropriated by the host state. Gas and oil pipelines are likewise potentially lucrative, but only if they are in friendly territory (Yarbrough and Yarbrough 1992; Hancock 2001), so that the nature of the relationship between two or more countries determines the value of the assets.

Economic integration can also lead to demands for economic policy changes in order to bring a state more into line with its economic partners. These additional policy changes, which make sense in a unified market even though they are not what a state would choose if left to its own political process, are costly to undo, so that if economic cooperation were to come to an end the state would be left with less-preferred market policies. In the lead-up to European Monetary Integration, for example, European economics changed their wage-setting institutions in order to take account of currency unification, sometimes in ways that ran counter to the interests of otherwise powerful groups (Enderlein 2006).

Much of the literature on security and interdependence recognizes, at least implicitly, the way in which political relationships between states influence the value of particular economic assets. Since trade requires specialization and specialization is costly to undo, firms have an interest in avoiding investments that will be likely to be disrupted by international conflict in the future. Uncooperative political relationships among states, then, reduce the potential value of economic investments (Morrow 1999). Joanne Gowa and Edward Mansfield (2004), for example, trace one implication of this by exploring inter- and intra-industry trade. They find that intra-industry trade, which results from economies of scale in specialized niches (and therefore attributable to firms’ conscious
investments in building up specific assets) is higher between states that have close security relationships and lower between states that do not. They reason that private actors will only invest in relationship-specific assets when they believe that the relationship will persist.

Suppose that states are unitary and that the same leaders who determine the future of the political relationship between states are also sensitive to the value of relationship-specific assets. If two states entering into a cooperative relationship both must make costly, relationship-specific investments, then they know that ending it would create extra losses for both of the states. Oliver Williamson shows that when external contract enforcement is uncertain or costly, mutual specificity of this sort can make contracts more stable (Williamson 1983), even in an uncertain environment (Riordan and Williamson 1985; Williamson 1994).

Policy coordination, though, may require one state to make a highly relationship-specific investment while a partner makes an investment that is much less specific, as when market integration leads one state (often the one with the smaller market) to make a larger adjustment than another. In choosing whether or not to enter into cooperation, the state with a higher potential level of specificity faces a dilemma: the vulnerable state (state V) has a valuable hostage at stake in the relationship, but its partner the non-vulnerable state (state N) does not.

Once states are already cooperating and one of the states is vulnerable to high adjustment costs if cooperation ends but another is not, the state with less at stake (state N) can use the threat of ending the relationship in order to take a larger share from V. Similarly, V (with more at stake) has a greater incentive to give up large concessions in order to preserve cooperation. If V’s outside option is poor enough – that is, if its potential losses from ending cooperation due to relationship-specific investments are great enough – it may be made worse off from a division of benefits than it would have been had it never started cooperating with N in the first place.

States that can be made worse off by entering into a cooperative arrangement in which they are likely to be exploited have a clear interest in refusing to cooperate at the outset. Instances of states actually entering into cooperation, and subsequently being exploited, should therefore be rare. Occasionally, though, either because of mistakes by leaders or because institutions sometimes produce poor decisions, states do enter into these unwise agreements. Thus, there are several examples that show this kind of exploitation can take the form either of explicit renegotiation of the terms of cooperation or of linking the continuation of cooperation to some other concession.

Prior to 1939, Germany used the implicit threat of ending economic ties with weaker Eastern European states to extract military commitments that these states otherwise would have been unlikely to make (Hirschman 1945). Relations between the United States and Hawaiian Kingdom are another example. Although an early trade reciprocity treaty provided Hawaii with a generous division of economic benefits, those benefits came at the cost of orienting the Hawaiian economy toward the U.S. mainland. As a result, Hawaii later agreed to cede Pearl Harbor in return for an extension of the agreement, a territorial concession the Kingdom’s leaders had previously resisted (Abdelal and Kirshner 1999). Some Canadians argued against ratifying the U.S.-Canada free trade agreement on the grounds that the gains from specialization would be dependent on continued access to U.S. markets. This could provide the U.S. with the
leverage, in the future, to renegotiate the terms of the agreement to be less favorable to
Canadian interests. Given the potential for future exploitation, it can be economically
eral rational for a country to reject a trade agreement even though it provides large joint
benefits and a generous division of gains in the present (McLaren 1997).
In more abstract terms, relationship-specific assets change the alternatives each
state has to cooperation, thereby changing the outcome of bargaining over the division of
benefits. Figure 2 illustrates this. Two states, V and N, are contemplating cooperation.
If they choose not to cooperate at all, they receive the payoff labeled “payoff to no
cooperation.” If they cooperate, they produce a joint gain labeled “payoff to
cooperation.” Cooperation, however, requires the states to invest in relationship-specific
assets, and requires V to invest more in such assets than N. So, if the states choose to
begin cooperating but then end their cooperation, they not only give up the benefits to
cooperation but they also each have to pay the costs of adjusting. When cooperation
ends, both states end up worse off than they would have been had they never started
cooperating in the first place – the point labeled “payoff to ending cooperation.”

Figure 2: Distributing joint gains with unequal relationship-specific investments

Note that V has more to lose than N (by assumption). As a consequence, when V
and N bargain over the division of gains from cooperation, V would be willing to accept
a deal that benefits N greatly while making V worse off than it would have been had it
never agreed to cooperate in the first place – the lower right end of the diagonal line –
since even this unpleasant bargain is better than ending cooperation entirely. If the states
divide the benefits from cooperation by splitting the difference as before, they end up
with a bargaining outcome that N would be happy with but would cause V to regret ever
having cooperated in the first place.
Anticipating just this sort of exploitation, many states in V’s position have declined to enter into cooperative arrangements that had the appearance of being generally beneficial, at least at first. The American decision to leave the British Empire was motivated in part by the desire to pursue a more self-sufficient path to economic development, since development within the Empire would lead to more dependence upon it, exposing the Americans to British extortion (de Figueiredo, Weingast and Rakove 2000). This rationale was repeated by postcolonial Indian leaders who rejected market integration and sought economic independence as a means to ensure political independence (Grieco 1982), by Latin American rulers who pursued import substitution policies during the Cold War, and Taiwanese leaders today who seek to diversify their economy’s ties away from the PRC. America’s Cold War allies – notably France – who sought to preserve military capabilities outside of NATO in order to reduce their dependence on the United States made a version of this calculation as well.

In other words, N has a credibility problem; it cannot commit, in advance, to a division of gains that would benefit both sides. This is as much a problem for N as it is for V, since it means that both states wind up with self-sufficiency, and thus are worse off than they could have been, in principle, from cooperation. These two states have a common interest in finding ways to solve N’s credibility problem, thereby making cooperation possible.

Contrived Symmetry through International and Federal Institutions

The problem is one of asymmetric vulnerability: cooperation can lead one state to invest in more relationship-specific assets than its partner, making the state more vulnerable than its partner. They can both be better off if there is some way for them to artificially make their outside options more symmetrical. When states try to resolve a commitment problem through “contrived symmetry,” they artificially manufacture a situation in which all states are vulnerable, rather than just one or a few of them, giving confidence to states that would otherwise be vulnerable to exploitation.6

States contrive symmetry by manufacturing something that they can all invest in together. Their goal is to reduce the extent to which vulnerable states lose bargaining leverage over time, since if they can reduce this loss in leverage they can make cooperation worthwhile for everyone and thereby make a negotiated solution possible. Since cooperation often necessarily requires states to invest in relationship-specific assets, it may not be possible for the states to contrive symmetry by reducing the extent to which vulnerable members make these kinds of investments. That leaves them with one alternative: they can create a relationship-specific investment for other members, including those that would otherwise be non-vulnerable, to contribute to. This investment, which all members forfeit if the cooperation ends, makes the bargaining among members more symmetrical.

---

6 The symmetry they create is “contrived” since it is something states “create in an artistic or ingenious manner” and “bring about by stratagem” (Merriam-Webster 2003).
While institutions influence politics in other ways as well, part of the power of institutions comes from adding to the costs that members must pay if and when they leave, beyond simply the direct costs of ending policy coordination. Political integration is not only costly to undergo, but it creates costs of secession that are distinct from those costs of leaving that arise from both the benefits of cooperation forgone as well as the costs that arise from lost investments in other relationship-specific assets. Institutions can make all their members vulnerable by leading all of them to invest something valuable in the institution. These investments can take the form of political parties, bureaucratic organizations, or possibly political identity. Federal constitutions are tools states use to contrive symmetry; so are international organizations, although to a (perhaps far) lesser extent.

In this section I describe these investments using a metaphor of posted bonds, and argue that political institutions have certain characteristics that make them ideal ways for states to post bonds. Institutions are relationship-specific, they can be linked to continued cooperation among states, and they can be costly for states to leave. Both federations and international organizations can have these characteristics.

The costs of leaving that stem from political institutions have a key characteristic: they are automatic in that they do not depend on any one actor (a federal government or IO secretariat) making an affirmative decision to levy some punishment. Such a threat to punish would not be inherently credible beforehand, since the very act of leaving an organization undermines that organization’s capacity and leverage in the first place. Bond-posting is an analogy commonly used to describe a mechanism that enacts punishment automatically. A bond – a payment that is forfeited if the payer reneges on a commitment – makes leaving costly, is valuable to political leaders in one state only as long as they preserve their relationship with another state.

Bonds have been studied extensively in other political contexts such as legislative bodies, as well as economic organizations such as firms (Williamson 1985) or “Better Business Bureau”-style associations (Kreps 1990). The posting of a bond is a useful metaphor for understanding the process of joining a federation as well, since a federal institution has many of the key features of a bond-holding mechanism when the governing coalition of a state is heavily invested in the political institutions of a federation.

---

7 Other ways in which institutions can alter outcomes include promoting transparency, creating common expectations, helping states to link issues and facilitate side-payments, and reducing transactions costs (Keohane and Martin 1995).
8 For example, members of the U.S. Congress are better off if they cooperate with each other, but they need assurances that they will honor agreements. Joining a legislative party requires an upfront investment (in time, money, or position-taking) that eventually pays off (in producing a record of legislative accomplishment, or electoral support). Because legislators’ reputations and policy positions are tied up with those of their parties, they are dependent on party organization and support. They have, in effect, built their careers around a party organization. This makes them vulnerable to censure from other members or party leaders, which in turn makes their promises to the party credible (Cox and McCubbins 1994).
Another key characteristic of the mechanisms that can hold bonds in a federation or international organization is that the cost of each state’s bond can differ. Once cooperation has begun, states will naturally attempt to lower their own exit costs while raising the exit costs of their partners, in order to create more bargaining leverage for themselves. To contrive symmetry, the states devise a system that causes them to post a bond with a central institution – a higher bond for the state (N) that would otherwise be less-vulnerable and a lower bond for the state (V) that would otherwise be more vulnerable. This has the effect of offsetting the bargaining disadvantage that the vulnerable state would otherwise have.

In the rest of this section, I propose three specific institutional mechanisms that can hold bonds among states and that meet the necessary criteria for contriving symmetry among states: political and legal organizations that adapt to specific rules and institutions, bureaucratic and military organizations, and (possibly, although not necessary to my argument) political ideology. I conclude the section by discussing the importance of how states bargain over control of an institution, and argue these three mechanisms states use to post bonds are also tools that states use to try to push the costs of exiting the institution on to each other.

**Parties and Leaders**

When a state joins a federation, it goes through a period of transition as people and organizations adjust to a new set of laws and institutions. The closer the union – that is, the more like a federal union and the less like an ad hoc agreement – the more the political system adjusts to the change. These changes, costly to undertake and then costly to undo, give politicians an incentive to maintain the new status quo. Although investments that the political parties make when they join a federation not be very costly (or costly to undo) when compared with the overall state, these costs are concentrated on the political leadership, so they affect the choices of those people whose decisions matter most.

Politicians and parties in a federation have an incentive to invest their scarce political resources in acquiring representation at the federal level. They also modify their political strategies at the state level to take account of politics at the interstate level. For example, American political parties at the state or district level pay some cost for being associated with a particular national (federal-level) party or faction when they national party affiliation leads them to defend positions taken by the party’s national platform that are unpopular in their own districts. This cost, however, is balanced against the benefits from being associated with broader party name recognition and record of governance (Ansolabehere, Snyder and Stewart 2001; Snyder and Ting 2002). If the federation ends, these local parties will be left with the worst of both worlds: policy positions inappropriate to winning locally, but without broader support from federal allies. Adopting a position suited to winning at the federal level is like an upfront investment: it costs, in terms of local support, but pays dividends in the long run since it allows a politician to steer benefits to supporters. This dividend, however, exists only as long as the federal institution survives.

The political adjustment costs that come with the end of an institution give parties incentives to support the status quo, even when underlying preferences in a state might point to secession. For example, even ethnically-based political parties in parts of Europe
where secessionist movements are pertinent use their leverage with their constituents to try to suppress the desire for autonomy or secession, because the leaders of those parties have a specific investment in the national political system as it exists (van Houten 2003). Separatists in Quebec were opposed by native Quebecoise politicians such as Jean Chrétien, whose larger concern once he became leader of the Liberals and Prime Minister was to maintain the relevance of the investments he and his party had made (Martin 2003). 9

By giving political parties much to lose if institutions fail, federal political systems create incentives for nearly any successful party or faction to support the continuation of the federal system. 10 In a federation, a party system that is national, rather than regional, is more effective at giving politicians incentives to maintain the federal system as a means to preserving their own careers. Even regional parties, though, can hold some investment in federal institutions as they are specialized to be effective at the federal level.

Bonds posted via political parties make leaving an institution – whether an international organization or a federation – more costly. They do not, of course, necessarily make leaving so prohibitively costly as to be impossible. States have little incentive to make exit catastrophically costly, for at least two reasons. First, political institutions are costly to join, and the cost of joining an institution rises with the level of political integration – federations are costlier to join than international organizations, which are in turn costlier to start up than ad hoc cooperation. So, states prefer not to create an institution that is impossibly binding on them, since that would be one that would entail more startup costs than are really necessary. Second, an institution from which it would be impossible to secede would not be in the states’ interests anyway, even in principle. The purpose of creating an institution is to contrive symmetry by making exit options worse. Once the outside options for everyone so bad that no one side has a large enough bargaining advantage over another to subject it to extortion, there is no reason to make those exit options even worse. The reason leaders create federal institutions is not to make exit impossible, but rather to make exit costly enough that the bargaining power among members does not erode asymmetrically. As long as political institutions make exit costly, they do not need to make it impossible in order to be effective.

Similarly, although it might be the case that the leaders of one state in a federation might be unable to fend off a bid from a rival group of potential leaders who wish to take their out of the federation (or IO), this merely reinforces the logic of my argument. Domestic competition within a state that is in a cooperative relationship (whether a federation, an IO, or some other form of cooperation) contributes to the costs to leaving for two reasons.

9 The outcome of the referendum, a majority in favor of continued union with the rest of Canada, may have hinged on the success of the national parties in campaigning against the measure (Dion 1996; Blanchard 1998).
10 “Big” decisions, such as whether to go to war, can be influenced by small political costs as long as those costs are narrowly focused on political leaders and there is no way for those leaders to avoid these costs. See, for example, the literature on audience costs (Fearon 1994).
First, deeply institutionalized cooperation between a state and its partners means that incumbent politicians within the state are tightly linked to that cooperation. The status quo creates winners, and those winners are willing to expend resources defending that status quo. From their perspective, leaving the federation is costly, and they impose costs on any of their domestic competitors who would seek to end the federation (or who would seek to bargain harder with other states in the federation, in a way that put the continued survival of the federal union at risk).

Second, if incumbent politicians stand to be disproportionately hurt if the federal union ends and, as a consequence, fight to defend it against domestic challengers, their survival will be increasingly tied to that of the federation itself. If the federation ends, then these politicians and parties will have lost not only the time and effort they invested in success at the federal level, but also the time and effort they invested in attempting to preserve the federal level itself. Since they will be made doubly worse off (and quite likely out of office) by a secession, the end of the union will be especially costly to them.

No classic federal constitutions (such as the American, Australian, or Canadian) specify secession procedures. Were there to be an actual secession in the Australian or Canadian federation, the outcome would be ambiguous and potentially risky for political leaders; it would certainly be costly (Clarke, Kornberg and Wearing 2000). Prior to the American Civil War, the legitimacy and procedure for secession from the Philadelphia constitution was unclear and in dispute, although most party leaders seemed to understand that secession would have been extremely disruptive to the existing political parties (Deudney 1995). The Civil War appears, initially, to be an example of federal institutions failing to prevent secession. An equally reasonable interpretation, though, is that American antebellum federal institutions worked surprisingly well, preventing a breakup of the union for over seventy years. Historians have typically argued that the second party system in the United States maintained confidence in the union by incorporating local politicians and governments into a national party structure that was invested in the national government (Holt 1978; Kornblith 2003). The national parties, recognizing their interest in preserving the union, simultaneously fought off regional rivals and formed coalitions between Northern and Southern interests. Both parties (Democrats and Whigs), for example, ran balanced tickets composed of pro- and anti-slavery candidates for president and vice president, and the Democrats, the dominant party, kept rules that gave Southerners an effective veto over national candidates and policies (Aldrich 1995, p. 133). Note that the nationalization of political parties prior to the war not only held the union together, it also protected Southern states from attempts by Northerners to renegotiate the terms of the union to abolish slavery.

The extent to which leaders suffer if their state leaves the organization depends on the nature of the institution – federations are costlier to leave than international

---

11 The literature on this point is large and detailed, but the general point is intuitive. Acemoglu and Robinson (2006) provide a recent example and demonstration. They show how incumbents are often able to prevent the adoption of new, social welfare-enhancing technology. If “credibly threatening secession” can be thought of as a “new technology” that a state can use to win benefits for itself, then their argument applies to my case as well.
organizations. Scholars studying particular IOs, such as the European Court of Justice or the European Court of Human Rights, have noted that these institutions can produce decisions that are difficult for states to undo or evade precisely because they work through commitments made by leaders within the state (in these situations involving international courts, they implicate commitments about legal procedures that courts within states have an interest in upholding). National leaders in Europe cannot ignore international legal rulings without undermining the legal basis of their authority at home (Alter 1998; Mattli and Slaughter 1998). At still lower levels of political integration, such as alliances or customs unions, the costs to leaders of exiting a bargain may be slight, but even here some scholars have argued that the institutions themselves can shape leaders’ calculation either because their domestic publics value commitments for their own sake (Fearon 1994), judge leaders’ competence by the extent to which they live up to commitments (Smith 1998), or value the process that leads to commitments (Lipson 2003).

Bureaucratic and military organizations

A second way states can post bonds is through the creation of specialized bureaucracies at the federal or international level. One particular example of such a specialized bureaucracy is a joint military force. Joining militaries has strategic implications that go beyond the merging and creation of other governmental organizations, but many of the initial consequences are the same. More than just functional cooperation, joining organizations merges them into a single administrative unit. Once agencies are merged, leaving cooperation is costlier for each state because it has to manage disentangling its administrative apparatus from that of its former partners. In one sense, this is little different than the bonds posted through political parties or other kinds of institutions that adapt to cooperation, but these agencies in general – and military forces in particular – have important strategic effects of their own.

Before considering these effects, however, first note that while military integration may be a means to make it costlier for states to exit an organization, and can therefore be a part of a federal agreement, military integration is not synonymous with federation. The American states, for example, kept their own militias for decades after the 1787 constitution, even though the constitution was federal in other respects. The constitution did not specify that the militias had to be merged into a single army for administrative purposes, and there was no uniform trend towards a national military even after the Civil War (Riker 1957). Similarly, the Australian states maintained separate standing armies for the first 14 years of federation, and only merged their militias as a

---

12 In some cases it is possible to imagine politicians even gaining, at least in the short run, from leaving an unpopular institution. In 2006, for example, the Iranian President contemplated ending his state’s cooperation with the International Atomic Energy Agency. Doing so would have been costly, since it would have ended security cooperation with the West and heightened suspicions about Iran’s nuclear program. These prospective costs were probably quite large, and stemmed from the end of cooperation itself. The specific institutional costs to leaving, on the other hand, were negligible; if anything Iran’s leaders may have benefited in domestic political competition by demonstrating their independence from international agreements.
result of World War One (Coulthald-Clark 1988, p. 121-123). Just as federation does not necessarily imply military integration, military integration does not necessarily imply federation. Most members of NATO participate in a unified military structure, in which there are officers of different nationalities within one direct military chain of command. In the 1950s, several Western European countries contemplated forming a European Defense Force, and Belgium, Germany, Luxembourg, and the Netherlands ratified a treaty that would have eliminated their national armies in favor of a European army that would have national divisions but be integrated at the corps level (Ruane 2000), and there are proposals today for a common European defense force even though the states have not formed a federal union.

The same is true for bureaucracies and agencies generally. The existence of specialized organizations is not itself a defining characteristic of federations, since for example the European Union has many such organizations whereas the United States did not, even relative to the state governments at the time, until the growth of the federal bureaucracy in the 1840s (Basinger 2003).

Bureaucratic integration raises the costs of exist for several reasons. First, the ability to regulate society is a basic characteristic of states. A state that merges its administrative apparatus with another state, and then later unmerges them, will be at least temporarily without the means to exert control. In the transition to reasserting control, it bears all of the administrative costs of taking control over state functions generally plus the additional costs and risks that come from momentary weakness or uncertainty.

Two more ways in which agencies can act as bonds are particular to military organizations. States can use their control of military organizations to raise the costs to other states of dissolving the union, either by using their military dominance to threaten punishment on a state that defects (for example, Russia’s use of its military against Chechen separatists) or by using their control of a military to resist punishment (for example, most American officers prior to the Civil War were Southerners, making it difficult for the Union, at first, to punish the secession attempt).

Finally, an integrated military may increase the costs to states of ending cooperation if the process of military integration leads to social integration. The Roman army, for example, was an instrument not just of coercion but of social engineering, since it brought together men from across the empire who then established a common Roman identity; some historians argue that the army was as important a tool of Romanization as any other (Goldsworthy 2003). Furthermore, integrated militaries that actually see combat may have an additional effect on both soldiers and citizens, solidifying bonds across as region – examples here include the Australian military in the first world war and the reunified American army in the Spanish-American War.

Identity and ideology

A third possible way in which states can post bonds is through a general change in the political identity of leaders or citizens. I introduce this possibility not because it forms a necessary part of the argument about how and why states post bonds, but rather to address an aspect of federal arrangements that, intuitively, might play an important role and to show that this role is entirely consistent with the theory I present. My argument so far has been explicitly rationalist, assuming that the leaders of states have had material goals. Suppose that, in addition, either leaders themselves or the citizens to
whom they are accountable are motivated by their identities as members of political communities.

Political communities are groups defined by the way members feel about each other. People identify with others in their community, believing themselves to be alike in fundamental and exclusive ways. A community is therefore characterized by a common “we-feeling” (Finnemore 2003). Once a common political identity has taken hold across states, individuals across those states are prepared to internalize the welfare of others within the community in return for recognition from them as legitimate members of the community.\(^{13}\)

In practice, how people define themselves and their community can vary. Any one individual has a variety of potential identities (such as Catalanian, Spaniard, Catholic, or European), and each of these suggests that the individual is part of a different larger community of similar individuals (Catalonia, Spain, Christendom, or Europe). Which of these identities is activated at any one particular time is at least partly a product of politics. In general, the historical and sociological evidence suggests that elites have the ability to influence, at least at the margins, which particular identity is most relevant to people. As a consequence, elites can influence the size and shape of a political community (Laitin 1986). Once constructed or activated, shared political identities can create constraints or imperatives for leaders, regardless of how those identities arose in the first place (Fearon and Laitin 2000). Most scholars, in any case, reject the idea that political identities are primordial (inherited from a distant past and not subject to change) and that shared political identity is often constructed by elites (Anderson 1983).

Whether shared identity in a community is a deliberate construction by political elites or an inevitable consequence of other institutional choices (such as military integration), it can constrain future leaders by making leaving a community more costly that it would be otherwise. Communities, then, whether in a federation or international organization – that is, whether we call them nationalisms or international communities – can be a tool for posting bonds. At the very least, the effect of political identity is that, all else equal, higher levels of political integration will lead to greater community or we-feeling, making exit more difficult.

Contrived symmetry

Federations and international organizations are tools that states use to contrive symmetry. That is, they raise the costs of cooperation ending, both to states that are

\(^{13}\) On recognition, see Alexander Wendt (1999; 2003). He describes moves toward mutual recognition within an ever-expanding community in teleological terms, as part of an inevitable process that will eventually lead to a global or near-global federation. Political leaders in such a world would have a much harder time seceding from organizations. Note that his argument leaves room for agency. Individual political choices, such as democratization in his terms, can lead to a greater or lesser feeling of community among states and their people. Political decisions about how to institutionalize cooperation may lead to a stronger or weaker affective community, resulting in more or less of a constraint on future choices. This is not entirely dissimilar from how Wendt’s argument might be applied to the development of a political community in Europe (1999, p. 297).
already vulnerable (state V) as well as to states that are not (state N). For these exit costs to have the effect of making the bargaining power among the states more symmetrical, however, political institutions need to impose higher exit costs on nonvulnerable states (N) than to vulnerable ones (V). If institutions impose the same exit costs on both kinds of states, then the difference in outside options favoring N is unchanged leaving the states with the same commitment problem as before, and if institutions impose higher exit costs on V than on N, they worsen an already unequal situation, making the commitment problem even more of a barrier to cooperation.

In federations in particular, states find ways to make it so that investments in institutions fall more heavily on nonvulnerable states than on others. Nonvulnerable states, of course, have a long-run incentive to allow this to happen. Unable to credibly commit to be restrained once its potentially-vulnerable partners are themselves locked in to cooperation, N is better off making itself more vulnerable, up to the point where its vulnerability is enough to give its partners confidence in their ability to bargain effectively within the federal institution they design).

Three examples show how institutions can be structured to create a higher bond for N than V. These are: ensuring that leaders from N are brought into political coalitions that separate them from the parochial interests of their constituents, tying N’s interests to specifically federal bureaucracies, and causing N to invest in physical assets that are specific to the federal institution such as a capital city in a federal district. I describe these in turn.

First, states can make federal institutions that are costlier for N’s leaders to leave than V’s leaders by the way they structure political competition at the federal level. Suppose that, in order to win power at the federal level, politicians have to strike a balance between the interests of their local constituents and those of other federal politicians (from different states) with whom they need to form a coalition in order to maintain power. To achieve this, the federal system needs to be set up in a way that gives leaders from state N an incentive to advocate policies that are contrary to those favored by their local constituencies. This can be done by establishing two conditions: first, leaders from N are likely to be in any governing coalition, and, second, leaders from N cannot govern the federation without drawing substantial support from other states with different interests.

Typically, vulnerable states in federations outnumber nonvulnerable ones. In Australia, for example, the leading vulnerable states (Victoria, South Australia, and Queensland) formed a bloc that acted more or less in unison in order to bargain effectively with the major nonvulnerable state, New South Wales; they also used their influence to exclude another nonvulnerable state, New Zealand, from the federation. Prior to federation, all parties understood that the leaders of the nonvulnerable state (new South Wales) would be key players in a governing coalition, but that the governing coalition would advocate far higher tariff rates than those leaders would have chosen had they remained in power solely in their home state.

Second, the institution can lead N to invest heavily in specifically-federal bureaucracies, creating more of a vested interest among civil servants and private agents who specialize in handling legal issues (lawyers, accountants, lobbyists, and so on) in state N than in state V. On the surface, this seems paradoxical, since stacking the federal bureaucracy with civil servants from state N should increase N’s influence in the
federation as a whole, rather than making N vulnerable. Having N penetrate the bureaucracy would increase N’s bargaining ability if the bureaucratic organizations in which it had disproportionate influence over divisive issues, such as allocating tax payments or distributing scare resources. However, if the bureaucracy N penetrates is one that provides public goods for the group of states as a whole with fewer distributive consequences, such as a military organization providing area defense or an agency regulating market rules, then this disproportionate investment can reduce the value of N’s outside option relative to V’s outside option. If the federation ends, N’s investment in agencies that produce regional public goods is either lost (in which case N has to pay high adjustment costs to revert to self-sufficiency) or persists (in which case N continues to provide the public good and V can free-ride off of it, bettering V’s outside option).

Third, the states can use the development of particular territory as a bond. Consider capital cities in special districts in federations, such as Canberra, Ottawa, and Washington.14 These are cities placed either near or within the territory of the nonvulnerable state – New South Wales, Ontario, and Virginia – and are created for the sole purpose of being the federal capital with little commerce or industry that does not specifically service the federal government. Upon joining a federation, the state that gets the capital city in or near its territory receives the economic benefits of a new city, but these benefits are tied to the preservation of the union and require costly adjustment if the union ends.

Bargaining over institutions

Cooperation, whether through a federation or international organization or without an institution at all, is something that states must jointly choose. Before specifying a hypothesis about the conditions in which states will choose one form of cooperation over either another form or no cooperation at all, it is necessary to specify the preferences that states have. This is particularly important here since federation is the result of a compromise in which at least one state does not get its most preferred outcome. Specifying the preferences of the relevant actors first is probably a good practice in any case (Frieden 1999).

Consider two states, V and N, contemplating forming a common market or providing for their common defense. (This can be a moment before the states begin cooperating, or, equivalently, a moment when these two states, already cooperating to some degree, are contemplating cooperating more deeply.) Assume that for these states the alternative to cooperation is self-sufficiency. They have a choice: should they avoid cooperation at all and pursue self-sufficiency? Should they cooperate through an ad hoc arrangement or through some low-cost institution like an international organization? Or should they form an institution that requires a large investment, a federation? V and N have different preferences because they each expect that the outcome they get will depend on which of these three options gets chosen. Assume, furthermore, that the states could both, in principle, benefit from cooperation, but N cannot credibly commit to a distribution of benefits that would make V better off by cooperating.

14 Ottawa is not in a distinct federal district, but is in a National Capital Region with a separate administrative authority.
Preferences

If there are no institutions to prevent it, vulnerable states that cooperate with nonvulnerable states risk being exploited. This is because V has high costs to leaving but N does not, so N is able to use its exit option, in effect, to renegotiate the terms of cooperation with V. In a federation, by contrast, all members face high costs to leaving, since federal unions contrive symmetry by creating vulnerability for states that would otherwise be nonvulnerable. The contrived exit costs, imposed on N through political institutions, make the bargaining situation more equal.

If given a choice between self-sufficiency and weakly institutionalized cooperation – either through an international organization or an ad hoc arrangement – V prefers self-sufficiency. V prefers federation to either of those. Since cooperation provides joint benefits, V would be happy cooperating as long as it could be guaranteed to receive enough of the share of the benefits from cooperation to make V better off by cooperating than not. Only a federation, which makes bargaining between N and V more symmetrical by imposing exit costs on N, achieves this.

So, states that would have high levels of relationship-specific assets in cooperation (state V) have federation as their first choice, self-sufficiency as their second choice, and international organization as their third choice. Federation is its first choice as long as the value to cooperation is high relative to the startup and sovereignty costs of federating.

The insight that actors in a more vulnerable position prefer a stronger institution is, of course, not new or novel. Economists studying the organization of firms, for example, have noted that workers and units within firms are more able to cooperate to their joint advantage when workers have some guarantee of getting a positive return from investments in specialization, and that this is more likely when intrafirm contracts are more complete or better enforced (Stole and Zwiebel 1996; Acemoglu, Antráas and Helpman 2005). Several political scientists have made similar points, such as Grieco (1982) and Smith (2000), who argue that states that have worse outside options (poor states in Grieco’s study, small states in Smith’s) have a preference for institutions that are more binding on their less-vulnerable partners.

Note that this preference ordering assumes that the gains from cooperation are high enough to offset the initial startup costs of forming a federation in the first place. To be precise, V has federation as its first choice as long as the startup costs are less than the value V stands to lose from having the weaker bargaining position that comes from being in an IO instead of the stronger bargaining position that comes from being in a federation. For any state to prefer federation to self-sufficiency, the joint gains from cooperation must be quite high. This would be the case when the form of cooperation at stake is a major issue, such as complete market integration or joint defense in a hostile world. So, federations are only likely to come about when the cooperation being contemplated is a big issue such as security or complete market integration. Issues that do not implicate a risk of survival or a major swing in prosperity are likely to be dealt with via an international organization or treaty – most, although not all, IOs deal with just one or a few issues, whereas federations are associated with comprehensive economic and security cooperation.
Now consider the less potentially-dependent state (state N). N is always better off cooperating through an international organization than a federation, for two reasons. First, once the relationship is underway it can get a better deal for itself through an IO that it would through a federation, since with an IO it has more leverage to renegotiate agreements later to extort better terms. Second, federations have higher startup costs than IOs. For state N, therefore, IO is always better than federation.

As long as there are any potential joint gains to cooperation (that is, the states working in concert can produce more than the sum of what they can produce separately), cooperation through no institution, or a weak institution such as an IO with a low contrived exit cost, is always better than self-sufficiency. So, international organization is the best possible outcomes from N’s perspective.

What is N’s second choice? If it cannot get cooperation through an ad hoc agreement or an international organization, N’s second choice will either be cooperation via a federation or no cooperation at all. Federation has some advantages: N will still be in a relatively favorable bargaining position within the federation and so will not be exploited, and N will get the benefits of cooperation. With self-sufficiency, N passes up the benefits of cooperation but avoids the start-up costs of federating.

How N ranks federation and self-sufficiency depends on the benefits of cooperation and on the startup costs of federation. If the startup costs outweigh any expected return from cooperation, then N prefers self-sufficiency second and federation third; otherwise N’s second choice is federation and third choice is self-sufficiency. The startup costs of federating, however, are themselves a product of choices that N and V make together. V, in order to protect itself from exploitation, prefers that N posts a high bond, but this of course makes federation less appealing for N. A bond that is high enough (where a high bond is a federal constitution) to make the difference, from V’s perspective, between cooperation being a bad idea and a good idea will also be high enough to make federation, all else equal, costly for N. So, for federation to be preferable to self-sufficiency, from N’s perspective, the joint benefits to cooperation must be high.

When the joint benefits to cooperation are high relative to the start-up costs of federation, states (like N) that would have low levels of relationship-specific assets in cooperation have international organization, as their first choice, federation as their second choice, and self-sufficiency as their third choice.

**Holding out**

If federation was the ideally-preferred outcome of each member of a group of states, then there would be nothing left to explain. The theory I am advancing, however, does not predict that there will ever be instances of this sort of harmony of interests. At least some states in a potentially cooperative group – those who would not automatically be made vulnerable via relationship-specific assets (like state N) – never have federation at their first choice, and always prefer cooperation through an ad hoc arrangement or, at most, an IO. No states in groups with similar levels of relationship-specificity (groups composed of all Ns or all Vs) would prefer federation, since there would be no need for institutions to contrive symmetry among states that were already symmetrical. Where federations emerge, they are always second-best for at least some of the members.

Here it is useful to distinguish between preferences and strategies. As Jeffry Frieden (1999) explains, actors have preferences over outcomes but they often are unable
to attain what they see as the best possible outcome. As a result, we often observe actors attempting to reach their second (or third, etc.) most-preferred outcome. Anticipating that they will fail if they push for their best outcome, they adopt a strategy aimed at securing an outcome that, although not ideal, is good enough.

N and V form a federation when V ideally prefers a federation and, if it cannot get a federation, will choose self-sufficiency over any other form of cooperation. Since cooperation requires the agreement of all parties, the ordering of V’s preferences gives it the ability to veto any form of cooperation other than its ideal. In other words, V holds out, denying cooperation unless it comes with a federal constitution.

N has a choice. It can propose its first choice, cooperation with a low level of institutionalization, such as an international organization; its second choice, federation; or its third choice, self-sufficiency. If N proposes its first choice of an IO, V will reject it, preferring that the states not cooperate at all. Both N and V, however, would rather cooperate through a federation than not at all. If N and V are able to negotiate a solution that leaves them each better off, then they together choose a federation.

To summarize: States will only form a federation when two conditions hold. The first condition is that the joint gains from cooperation are valuable enough to outweigh the startup costs of federating. The second condition is that cooperation requires some states to make a disproportionately valuable investment in relationship-specific assets.

This argument for the origins of federations can explain not just why states form federations instead of pursue self-sufficiency but also why they form federations instead of simple alliances or customs unions. Only a federation, which creates high political costs to exit as a substitute for other exit costs, can rebalance bargaining and make cooperation mutually appealing.

In broader terms, states that would be put at risk from extortion once they are in a cooperative relationship demand political integration from their partners as a precondition for cooperation. To avoid being subject to extortion themselves, they turn the extortion around by demanding as the price of cooperation, upfront, political integration that is as deep as they can get from their partners. Their partners give in when the benefits of cooperation are great enough that they are willing to subordinate themselves to political integration, thereby getting otherwise unattainable cooperation.

Australia’s Experiments with International Organization and Federation

Why is Australia one country instead of six? Why are Australia and New Zealand two different countries? Although in retrospect Australia and New Zealand seem to have logical boundaries, there was nothing inevitable about them at the time. The inability of New South Wales and Victoria to reach an agreement on a customs union via an international organization, but then the success with which they negotiated a federal union, is superficially puzzling. Thinking of a federal constitution as a device to contrive

---

15 Even Germany, formed when Prussia threatened its partners with conquest if they did not voluntarily join, was voluntary in that it resulted from decisions by independent states. The small states simply negotiated themselves out of existence (Fazal 2002).
16 That is, I assume that there are no other impediments to bargaining, such as incomplete information or sources of mistrust other than the ones I have already taken into account.
symmetry, however, resolves the puzzle, since the root of the problem with Australian proposals for an international organization in the late 19th century was the inability of New South Wales to commit to refrain from revisiting the terms of cooperation.

This section proceeds begins with an overview of the political process that resulted in federation. Second, I describe the bargaining between New South Wales and Victoria. Third, I explore the federal institution itself, as proposed in the 1890s, to show how it held bonds from the states and how it was able to extract a higher bond from New South Wales than the other states; in the course of this, I also account for the struggle between New South Wales and Victoria over how to approach New Zealand’s controversial candidacy for membership.

Overview

Beginning with New Zealand in 1853 the Australasian states – New South Wales, New Zealand, Queensland, South Australia, Tasmania, Victoria, and Western Australia – became self-governing dominions within the British Empire when the British government began to pursue a policy of disentangling itself from its settler colonies. By 1860, when Queensland became self-governing, all the states except the nearly empty Western Australia were governed by locally-elected parliaments. From the 1870s, other states such as Germany and the U.S. pursued their interests with the colonies “on the basis of de facto, if not de jure equality of status” (Veit-Brause 1988, p. 143). Map 1 shows the states with the years of self government.

Map 1: Self government in the Australasian states (distances are to scale)
Most infrastructure was local, as the states built railroads connecting the farms, pastures, and mines of the interior to the major port cities; there was little demand for transportation links between the major cities, which were arrayed along the coast (Clark 1908; Harman 1970). Still, the states had several common goals, and newspapers and political memoirs emphasized the importance of regional cooperation. At an inter-colonial conference in 1863, the states rejected a proposed formal international organization and instead instituted a regularized system of informal meetings. Decisions made at these meetings did not legally bind the participating colonies – the agreements the delegates reached often went unimplemented. The conferences considered everything from divisive issues like tariffs to more mundane coordination issues like marriage regulations, agricultural standards, common postal rates and routes, and telegraph lines. In general, they did not reach or implement agreements on trade (Moore 1910; Livingston 1932; Mines 1969; Dalziel 1987). In practice the states had relatively open trade with each other through the 1860s, but by the end of that decade they began to erect substantial trade barriers against each others’ products.

In 1881 the premier of New South Wales, Henry Parkes, proposed an IO system incorporating all the self-governing Australasian colonies (which included New Zealand). The proposal was for a legislature with the power make decisions on behalf of the group, with no administrative apparatus to carry out decisions (Wise 1909, p. 164). The other states were not interested and pushed for an alternative plan with a more centralized government. Eventually both proposals were abandoned and several of the states, Victoria, Tasmania, and Queensland, began formally cooperating through another IO they called the Federal Council (not really a federal system, its institutions resembled the contemporary EU although it did not include a proposal for a customs union). New South Wales and New Zealand both participated in the negotiations over the formation of the Council, but both states declined to join, New Zealand because there seemed limited benefits, and New South Wales since its leaders seemed to feel they could hold out for a better deal that would give it the leeway to pursue its own trade policy (Beach 1899). In particular, chambers of commerce and the allied Free-trade party in New South Wales were always skeptical of the value of a strong central authority (Turner 1904, p. 332).

Throughout the 1880s, Australian markets became more and more divided from each other as protectionist parties tightened their grip in Victoria and the other smaller colonies. New South Wales and New Zealand remained open to trade; while their interests were harmed by lack of regional market access, they still benefited from trading on world markets. Still, the lost opportunities for regionalism eventually led the government of New South Wales in 1889 to agree to open federal negotiations. A series of conventions met on and off for the next decade, and the states finally ratified an agreement in 1899, and the federation took effect on January 1, 1901.

Australians studying their origins have long noted that the relevant question to ask is why the Australian states federated when they had an alternative available: coordinated action through decentralized cooperative arrangements, such as a customs union or an alliance (Baker 1897; Reeves 1902; Irving 1999; Hirst 2000). To explain Australian federation, a theory needs to account not just for why the states federated, but also why they did not instead choose to cooperate via a system of customs unions and military
alliances as they had originally discussed, why each of the six states joined but New Zealand did not, and also why they did not federate sooner.

Although the states formed a federation in 1901, interested observers at the time certainly did not see it as inevitable and many predicted that it would fail (Parker 1964, p. 156; Pringle 1975). Writing in his memoirs, the attorney general of New South Wales revealed that he had had grave doubts about the prospects of federation. “There was, too, such a sullen and deep-seated hostility between New South Wales and Victoria, that at times it seemed as if the history of the South American republics might repeat itself” (Wise 1909, p. 162). The drafters specifically considered retaining weak cooperation through an international organization as an alternative to the federation, and in each state there were at least some political leaders advocating this course as an alternative to federation (Beach 1899).

Federation might have included New Zealand. New Zealand had been settled by the same people, had trade ties with the Australian mainland, and participated in regional meetings and the regular premiers’ conferences. Geography alone is an insufficient explanation; Tasmania, also an island, joined the federation, and Western Australia joined even though Perth (its major city) was further from Sydney than Auckland (New Zealand’s major port) and could only be reached by sea (Grey 1901, p. 215). Keith Sinclair, New Zealand’s foremost national historian, argues forcefully that the colonial histories of New Zealand and the other Australasian colonies clearly show that separation was not inevitable, and that New Zealand and Australia did not have “separate manifest national destinies.” “No one contemplating inter-colonial negotiations in the eighteen-seventies would have had convincing reason to forecast that within thirty years six of the colonies would federate. It would have been equally difficult to foresee that New Zealand would stay out” (Sinclair 1970, p. 153 and 175).

Australia is in many ways an ideal case for exploring the logic of my argument. Within one regional group of states with similar cultural origins and political-legal systems there is both variation in the independent variables of interest as well as the dependent variables. The process the states went through to federate was open enough that it is possible to test several implications of the theory. Australia is also a paradigmatic case of federation – along with Canada, Switzerland, and the United States, Australia is often cited as one of the most unambiguously clear voluntary federation in history (Wheare 1946).

New South Wales and Victoria

Federations form when potentially vulnerable states value cooperation but perceive that cooperation will result in them investing proportionally more than their larger partners into relationship-specific assets. In Australia, in the second half of the 19th century, Victoria was in exactly this position with respect to its more populous neighbor, New South Wales.

Initially (the 1850s and 60s), the Australian states all exported primary goods, agriculture and minerals, to Europe and North America. The states were relatively open to trade with one another, with economic policy being coordinated through ad hoc conferences. Despite this openness, trade was at relatively low levels since their economies were not specialized and dealt mostly in overseas markets. New South Wales (home of Sydney) was the largest and wealthiest, and made its money exporting grain
and wool. It had a weak preference for openness and cooperation, and for self-sufficiency (at least with respect to the other Australian states) if it could not get regional openness. Its parliament continually expressed disinterest in any serious proposals for federation or any other form of political integration (Wise 1909).

Victoria (home of Melbourne) likewise exported primary goods, and preferred this tacit cooperation, at least at first – within the state there was little interest in federation during the first decade of self-rule (Garran 1933). After Victoria’s 1864 election, though, its position began to change. Starting then, its governments pursued protection for industrial development (Turner 1904, p. 114-15); this protectionist coalition cemented its hold on power in 1872 and variants of it held power in the state through 1901 (Macfie 1893). New South Wales, in contrast, maintained a steady free trade policy throughout, with only minor interruptions of rule by a protectionist coalition (Hiscox 2002, p. 119). Likewise, New Zealand governments consistently proclaimed a free trade policy (Douglas 1909, p. 194-99). The other states fell between these extremes, with South Australia an intermediate case (Reeves 1899; Clark 1908).

Chart 1 illustrates the difference between New South Wales and Victoria. The vertical axis is the average tariff rate, defined as a ratio of customs revenue over the total value of imports. The solid line is the tariff rate for Victoria, the dashed line is for New South Wales. Elections produced a short-lived protectionist coalition in New South Wales in the early 1890s, but at other times was the state was consistently more open to trade than Victoria, and the difference only increased over time.

Chart 1: Average tariffs rates in Victoria (solid line) and New South Wales (dashed).
Melbourne was industrializing behind tariff walls, Sydney (in NSW) was remaining primarily a port city for agricultural trade. If Victoria could have access to New South Wales’ market, it could expand its industrial production and grow rich.

Market integration would only be profitable for Victoria, however, if the common market were protected from imports of manufactured goods from Europe. If New South Wales opened its market and agreed to a customs union, and then Victorians invested in an industrial expansion, and then New South Wales left the customs union, Melbourne would be saddled with an enormous overcapacity of uncompetitive industries. This position of vulnerability frightened Victorian leaders; in their memoirs (Westgarth 1887; Duffy 1890) and public speeches (Willoughby 1891; Macfie 1893) they often referred to the dilemma they faced: on the one hand, they hoped to promote industrial development, but on the other hand, they felt New South Wales was an untrustworthy partner.\footnote{This is by no means unique to Australia. Members of contemporary regional preferential trade agreements are often motivated by the impulse to provide a large, captive market for infant industries. Mercosur, in Latin America, was at least partly a tool to protect regional manufacturers (such as Brazil’s construction equipment producers) from competition from more industrialized countries by granting them privileged access to a regional market (Mattli 1999), and some of the appeal to France of European integration in the 1950s was that it would give its farmers a larger, captive market (Moravcsik 1998). The point here is that these captive markets actually have to be captive (without an easy exit option) in order to have the desired effect.}

Negotiations over a customs union between Victoria and New South Wales failed, even though the local chambers of commerce in both Sydney and Melbourne supported a common market (Hirst 2000). The governing coalition of New South Wales was led by the Free Trade party, whose very identity (as the name implies) was that it promoted open access to world markets. Henry Parkes (leader of the Free- Traders and premier for much of the 1880s) adopted as his government’s negotiating position that New South Wales would not accept an agreement that restricted its freedom of action, so that any common external tariff that the common market adopted as part of the original negotiations would be subject to a downward revision later. He viewed this as a reasonable stance, since any potential revision would be voluntary by all parties, and a state could secede from the agreement if it needed to (Duckworth 1899; Galloway 1899; Allin 1907; Clark 1908).

This was Victoria’s dilemma. Once invested in a common market, Melbourne naturally would agree to any revision Sydney proposed if the alternative was for New South Wales to pull out of the agreement. A common market run by New South Wales’ terms, once Victorian industries had already adapted to the common market, would be worse for Melbourne’s manufacturing cartels than never joining a common market in the first place. Melbourne’s industrial elites, for example, observed the American Civil War with interest. Watching the Northeastern cities lose, and then regain, access to the Southern agricultural regions made them realize their own potential vulnerability to a break in their growing continental market (Moore 1970). The alternative was to push for a federal constitution that would prevent Sydney from even contemplating leaving (Serle 1969).

While the leaders of all the states in the region expressed support, in principle, for an agreement that would provide open market access, none were actually negotiated. In
1883 Tasmania and Victoria proposed and signed a full customs union, but protectionist opposition in Victoria led the government to drop it. Victoria and Queensland started to negotiate a treaty in 1886 but set it aside, and South Australia and Queensland tried but failed to reach an agreement. These were minor failures, in any case: the real prize for Victoria and the smaller states was the elusive goal of getting New South Wales to commit. In 1888, New South Wales and Victoria brought decades of fruitless negotiations to an end when they failed to reach any trade agreement.

For Victorians, the danger of a customs union was that New South Wales might later change the terms of the agreement. The value of a customs union to Victoria was that a common market under high protective tariffs would make New South Wales a market for Melbourne’s new manufactured goods. If, however, New South Wales later decided to leave the agreement in some way – either by formally seceding or by informally dropping its trade barriers against European and North American goods – Victorian goods would no longer be competitive and Melbourne would effectively lose its captive market. As these tariff agreement proposals failed, the states continued escalating tariffs on each others’ products, so that by the 1880s they were in what appeared at the time to be a trade war (Westgarth 1887; Galloway 1899; Wise 1909; Allin and Anderson 1929).

The leaders of Victoria, as well as the other smaller states in the region that had been pushing federation (chiefly South Australia and Tasmania) understood that having a federal constitution would go a long way to resolving their fears about the federation. A common phrasing, made famous by pro-federation leaders in South Australia, captures the sentiment. South Australia was a state that, like Victoria, would have made large relationship-specific investments in the event of market integration by reallocating labor from wheat production to more specialized agriculture (like grapes). A newspaper there captured the unique appeal of federation over a customs union when it explained that the federation would bring about a common market independent of the “good will or caprice” of the other states. Federation, which would make it more difficult for New South Wales to change course, provided insurance as a way states could get a common market they could trust (Norris 1975, p. 18-19, 28).

Although Victoria preferred federation to self-sufficiency to IO, prior to 1889 New South Wales preferred IO to self-sufficiency to federation, and no deal was reached. The situation seemed to change in 1889, when Henry Parkes, the leader of the free trade coalition in New South Wales, relented and agreed to a federal constitutional convention. His governing coalition promptly fell apart, however, when he was abandoned by Sydney financiers who served as intermediaries in trade between New South Wales and Europe and therefore benefited from global trade and would therefore be harmed by a high common external tariff. There followed a three-year period of rule in New South Wales by a protectionist coalition that took advantage of the split within the free-traders (Knightly 2000, p. 55; Travers 2000).

Once the free trade coalition returned to power in Sydney, and following another period of internal debate, it advocated federation despite misgivings. In the interim, the growth of potential markets in Australia began to make access to those markets nearly as appealing as the overseas markets New South Wales would lose by federating. The free-trade leaders all would have ideally preferred a customs union, but noted that
There are some people who suggest that we can have Intercolonial Freetrade without Federation; they must be very sanguine and unpractical. Ever since 1873... attempts have been made by the various colonies to enter into customs treaties with each other, &c., &c., all of which have ended in entire and dismal failure, although they were confined to a few articles of colonial production (Baker 1897, p. 7).

Given the lack of an alternative, private groups such as the New South Wales chamber of commerce, reluctantly changed their views and backed federation (Martin 1964, p. 223).

*New South Wales, New Zealand, and the structure of federation*

New South Wales and New Zealand were in similar situations. By the middle of the 1880s, both states had effectively established lucrative markets for exported primary products. For New Zealand in particular, this development had been surprising. Early on after self-government, New Zealand was as economically integrated with the other states as New South Wales was. Later, however, the New Zealand economy gradually diverged from the others through the 1870s. Trade across the Tasman Sea (between New Zealand and the East coast of Australia) did not grow as fast as commerce with Europe.

In 1882, some shipping lines introduced refrigerated freight transport. This allowed New Zealand to export meat directly to Europe’s growing middle class, and revitalized the state’s pastoral regions. Even if the Australian market were to open completely to it, New Zealand would not be able to engage in any production that more effectively capitalized on its comparative advantage (Galloway 1899; Sinclair 1970). The magnitude of New Zealand’s trade with Britain following the first frozen meat shipment in 1882 was staggering: in one year alone, 1898, New Zealand exporters sent 2,700,000 frozen sheep carcasses to Britain, representing two-thirds of total British mutton imports. Meat exports were more important to the New Zealand economy than simply the primary production, since the trade generated demand for other industries. Much of New Zealand’s economic growth over the two decades following 1882 was popularly attributed to related developments as new firms made investments in freezing works, communications networks with British and other European markets, and shipping and dry-docks for long-distance merchant ships, creating a fundamental shift in the New Zealand economy (Irvine and Alpers 1902, p. 364-8).

By the 1890s New Zealand sent a lower percentage of its exports to the other Australasian colonies than any of the others did to each other. Still, federation was controversial. Although ranchers and the new industries that served long-distance trade opposed federation, there was support for federation among grain farmers and exporters of forest products who could have benefited in principle from trade with arid Australia and who feared being excluded from the Australian market (Sinclair 1986; Sinclair 1987). The strong balance of interests, however, favored staying out (Reeves 1901).
Chart 2 shows this graphically, by plotting the total regional trade (exports and imports to and from the six other Australasian states) as a proportion of all trade (all exports and imports). In the chart, New South Wales is the solid line at the top, Victoria is the dotted line that gradually rises, and New Zealand is the dashed line that falls dramatically.

Within both New South Wales and New Zealand, light manufacturing industries had supported protectionist policies in the past. These groups now opposed federation, since federation would result in increased competition for the home market with manufacturers from Victoria. In New South Wales, a strong coalition of agricultural interests that hoped to increase regional exports supported federation, since it would give them secure access to regional markets. These groups traditionally supported free trade. Other traditional free-traders in New South Wales, though, included groups that depended on exports to Europe. These tended to oppose federation, at least before the mid-1890s. Although free trade groups in New South Wales were divided, in New Zealand, in contrast, most of the traditional free-traders were in the business of exporting to Europe. So, in New Zealand both protectionists and most free-traders opposed federation.

In that sense, New Zealand illustrates a counterfactual case to New South Wales. Had Sydney not moved to a position where the benefits from regional integration outweighed the potential losses from losing autonomy (and had Sydney not recognized that it was never going to get economic integration on its own terms, as an IO), the relationship between NSW and the rest of Australia would very likely have unfolded just
as the relationship between New Zealand and Australia has — periodic economic cooperation, but political independence.

At the early constitutional conventions in the beginning of the 1890s, when New Zealand was a full participant in the negotiations and most participants still believed New Zealand was seriously considering joining the federation, there was a conflict over the terms of New Zealand’s potential entry. New South Wales hoped to persuade New Zealand to join the federation while Victoria sought to put in place procedural hurdles that would make it more difficult for New Zealand to join, and in particular Victoria sought to make it difficult for New Zealand to join the federation at a later date, after the original members had already set up a government.  

Conventional wisdom predicted that in the hypothetical federal parliament New Zealand’s representatives, like New South Wales’, would be more inclined to favor open trade with the rest of the world than representatives from the other states (Reeves 1899; Wise 1909). Without New Zealand in the federal parliament, Sydney’s representatives would be in a distinct minority, and would either have to endorse substantially higher tariffs than they would otherwise prefer or be relegated to being in the opposition (Griffith 1896; Robertson 1897; Barton 1901; Allin 1918).

Victoria’s strategy can be interpreted in two ways. First, more straightforwardly, it was a way of keeping Sydney’s free-trade preferences from gaining too much traction in a then-hypothetical future Australian parliament. If this had been Victoria’s only goal, however, then it is puzzling that it went to the effort to take on this contentious issue when it could have advanced the same end of reducing New South Wales over all influence in other ways. For example, throughout the 1890-91 conventions there was very little disagreement about the composition of the lower house of parliament — seats would be apportioned to the states on the basis of population, and New South Wales had the most residents. In principle, though, seats could have been distributed in a way that reduced New South Wales’ advantage by making fixed apportionments (Baker 1891). Alternatively, Victoria could have made efforts to reduce the influence of the lower house, apportioned by population in which New South Wales had an advantage, and increase the influence of the upper house, with six senators from each state. Instead, Victoria did not push either of these changes.

The second interpretation of Victoria’s strategy was that it was, in fact, happy to let New South Wales send all of its ablest politicians to the Australian parliament, and to let them play an influential role in making policy for the federation. Getting Sydney’s political elite invested in a national government was, after all, the entire point. Victoria’s aim was simply to prevent any chance of the emergence of a free-trade majority coalition.

So, in contrast to the Federal Council which gave equal weight to each of the member states, the federation gave greater representation to the larger states by creating two houses of parliament. The drafters put a great deal of effort into making it so that the Australian executive would be accountable to both houses of parliament (Baker 1897).

---

18 Sydney sought to include a constitutional provision that New Zealand could join at a later date on the same terms as the other states, e.g. with seats in the two houses of parliament apportioned in the same way. Victoria won, instead, a constitutional provision that specified that any new state would be admitted only on terms agree to by the parliament, implying that it might not be permitted a many representatives.
When voters in New South Wales rejected the first constitution and Reid reopened negotiations over the constitution, representatives from Victoria quickly agreed to a proposal to increase the influence of the lower house at the expense of the upper house; this change had the largest proportional beneficial effect on New South Wales. Specifically, the changes gave some precedence to the lower house in the case of a deadlock between the lower and upper houses. In constitutional and legal thought at the time, theorists placed a great deal of importance on the tie-breaking procedure, since the framers anticipated that differing modes of election would lead to frequent disagreements between the two legislatures (Beach 1899; Warden 1992, p. 151).

The other change to the constitution was to specify, for the first time, the location of the federal capital: it would be in a federal district within the territory of New South Wales, but at least 100 miles from Sydney. Reid expected the capital city to provide some economic benefit and was careful to hint that the site chosen would be to the south of Sydney, near the border with Victoria. The pastoral towns in this region otherwise opposed federation, but the possibility of being bribed with the capital may have swung some referendum votes (Hewett 1969).

In order to join the federation, Victoria and the other smaller states had to adapt economically. New South Wales, however, more than any of the other states, had to adapt politically. Its leading politicians, unlike those from any other state, had to pivot upon entering national politics and endorse a protectionist tariff. For example, George Reid, a committed free-trader in the Gladstonian tradition while premier of New South Wales, quickly became a fervent protectionist once in the Australian parliament – a change he justified as a response to the new reality that federation had “settled” the tariff question (Reid 1917) but that coincidentally allowed him to be Prime Minister of the federation after a decent interval (Crisp and Hart 1990). As expected, the federal capital city (eventually named Canberra) created an economic boom for the nearby towns to the south of Sydney, but the capital region generally lacked any other industry and to this day remains dependent on the presence of the capital for its economic activity.

Federation, then, was a tool for getting New South Wales to commit to economic cooperation. Absent political institutions, it would have easily agreed to a customs union but would never have been able to refrain from pursuing its own economic interests by reopening to world markets. By tying Sydney’s leaders to Australian parliamentary institutions, Victoria managed to tie New South Wales to Australia.

Conclusion

Federation is a costly tool that the leaders of states use in order to achieve gains from cooperation that would otherwise be impossible. Faced with the risk of being exploited in an international organization, potentially-vulnerable states are sometimes able to extort federal commitments from less-potentially-vulnerable states as the price of cooperation.

---

19 At the time of the drafting, the framers anticipated that deadlocks would be common (Barton 1901; Reeves 1902), but as it happened they have been quite rare (one in 1975 nearly caused a constitutional crisis, but there had been few previous incidents).
References


Allin, C. D. and W. Anderson (1929). Australasian preferential tariffs and imperial free trade a chapter in the fiscal emancipation of the colonies. Minneapolis,


Baker, R. (1891). A manual of reference to authorities for the use of the members of the National Australasian Convention, which will assemble at Sydney on March 2, 1891, for the purpose of drafting a constitution for the Dominion of Australia. Adelaide [S. Aust.], W.K. Thomas & Co.


Baker, R. (1897). Federation. [Adelaide, s.n.].


Smith, A. (1776). *An inquiry into the nature and causes of the wealth of nations.* London, Printed for W. Strahan; and T. Cadell ...


