

NATIONAL SOVEREIGNTY IN AN INTERDEPENDENT WORLD*

Kyle Bagwell[†] Robert W. Staiger[‡]

November 2017

Abstract

What are the sovereign rights of nations in an interdependent world, and to what extent do these rights stand in the way of achieving internationally efficient outcomes? These two questions rest at the heart of contemporary debate over the role and design of international institutions as well as growing tension between globalization and the preservation of national sovereignty. In this chapter, we propose answers to these two questions. We do so by first developing formal definitions of national sovereignty that build on features of sovereignty emphasized in the international political economy literature. We then utilize these definitions to describe the degree and nature of national sovereignty possessed by countries in a benchmark (Nash) world in which there exist no international agreements of any kind. And with national sovereignty characterized in this benchmark world, we then evaluate the extent to which national sovereignty is compromised by international agreements with specific design features. In this way, we delineate the degree of tension between national sovereignty and international objectives and describe how that tension can be minimized – and sometimes in principle even eliminated – through careful institutional design.

*This is a revised and updated version of our working paper Bagwell and Staiger (2004). For many helpful comments on the previous version, we thank Robert Keohane, Stephen Krasner, Alberto Martin, Jon Pevehouse, Jeremy Rabkin, Donald Regan, Guido Tabellini and seminar participants at Berkeley, Notre Dame, the Stockholm School and the NOITS Workshop in Copenhagen, and we thank the National Science Foundation (award number SES-0214021) for support. We are also grateful to Lili Yan Ing and Miaojie Yu for providing very helpful comments on this version.

[†]Department of Economics, Stanford University; and NBER.

[‡]Department of Economics, Dartmouth College; and NBER.

“Of all the rights possessed by a nation, that of sovereignty is doubtless the most important.” Emmerich de Vattel in *The Law of Nations*, as quoted in Jeremy Rabkin, *Why Sovereignty Matters*, p. 27.

1. Introduction

What are the sovereign rights of nations in an interdependent world, and to what extent do these rights stand in the way of achieving internationally efficient outcomes? These questions rest at the heart of contemporary debate over the role and design of international institutions as well as growing tension between globalization and the preservation of national sovereignty. But answers are elusive. This is attributable in part to the fact that national sovereignty is a complex notion, reflecting a number of different features. And it is attributable as well to the fact that nations interact in increasingly complex and interdependent ways, making it difficult to draw clear distinctions between international and domestic affairs.

In this chapter, we propose answers to these questions. We do so by first developing formal definitions of national sovereignty that build on features of sovereignty emphasized in the international political economy literature. We then utilize these definitions to describe the degree and nature of national sovereignty possessed by countries in a benchmark (Nash) world in which there exist no international agreements of any kind. And with national sovereignty characterized in this benchmark world, we evaluate the extent to which national sovereignty is violated by international agreements with specific design features. In this way, we delineate the degree of tension between national sovereignty and international objectives and describe how that tension can be minimized – and sometimes in principle even eliminated – through careful institutional design.

We begin by describing a benchmark two-country model of international interdependence. In this benchmark model, an international “externality” variable defines the interdependence between the two countries, and this variable is modeled in a way that is general enough to allow the nature of this interdependence to take a variety of possible forms, ranging from international trade to the depletion of a common-pool resource to global climate change. Within this benchmark model, we develop a working definition of sovereignty.

Our starting point for defining sovereignty is the Westphalian norm of “non-intervention in the internal affairs of other states” (Krasner, 1999, p. 20). To make this norm operational, we

must define “non-intervention” and “internal affairs.” Building on the notions of sovereignty commonly discussed in the international political economy literature, we combine elements of several of these notions and say that a governmental decision problem concerns the country’s internal affairs whenever its payoff in that decision problem does not depend on the actions of any extraterritorial agents; otherwise, we say that this decision problem concerns the country’s “external affairs.” In the context of voluntary international agreements, we then say that one state has intervened in the internal affairs of another state, and therefore that a violation of sovereignty has occurred, whenever an international agreement leads a country to make commitments over matters that concern its internal affairs or to make commitments that alter the normal operations of its domestic institutions within the domain of its internal affairs. We argue that our formal definition of sovereignty captures three key features of Westphalian sovereignty emphasized in the international political economy literature that seem especially relevant in the context of voluntary international agreements: first, commitments that result from voluntary international agreements do not necessarily violate Westphalian sovereignty (as when these commitments pertain only to a country’s external affairs); second, international commitments over policies that concern “sufficiently domestic” affairs (i.e., internal affairs) *do* violate Westphalian sovereignty; and third, international commitments that distort the normal operation of domestic institutions also violate Westphalian sovereignty.

With our formal definition of sovereignty in hand, we then turn to a characterization of the nature and degree of sovereignty that countries possess according to this definition in various economic environments and institutional settings. We begin this characterization within our two-country benchmark model. To identify the degree of sovereignty that countries possess in this environment in the absence of an international agreement, we show that a government’s policy choices in the Nash equilibrium can be partitioned into two choice problems: a choice of the externality variable and the contribution that this country makes to the determination of the externality variable, given the other government’s policies; and a choice of how best to use its policy instruments to achieve its objectives while delivering the given contribution level. With this partition, we are able to associate a country’s external affairs with the first choice problem, and its internal affairs with the second. Intuitively, when countries are mutually interdependent, the external affairs of each country consist of that country’s choices over its contribution to the determination of the externality variable and the equilibrium level of the externality variable, since its payoff in this choice problem depends on the actions of external actors. By contrast,

the matters that concern the internal affairs of each country are that country's choices among all its policy combinations that are consistent with a given contribution to and level of the externality variable, since its payoff in this choice problem is independent of the actions of external actors.

We put our definition of sovereignty to work by evaluating according to this definition the consequences of various forms of international agreements for national sovereignty. Specifically, we consider first whether it is possible to eliminate the inefficiencies that arise in the Nash equilibrium with international agreements that are limited only to the external affairs of each country, and thereby to navigate all the way to the international efficiency frontier without violating national sovereignty by means of such agreements. Our first main result is that this is indeed possible within the benchmark model. That is, we show that it is always possible within the benchmark model to pick any point on the international efficiency frontier that could be achieved by international negotiations over all policy instruments, and to achieve that point with international negotiations that are limited only to the external affairs of each country, i.e., to the level of the externality variable and each country's contribution to it.

We next consider the way in which a country's sovereignty is violated within our benchmark model when the country negotiates international commitments that concern its internal affairs. Such commitments directly violate a country's sovereignty, but we show that direct violations of sovereignty can also imply further indirect violations of sovereignty as well, under which government decisions that are not the subject of international negotiation are nevertheless distorted away from the decisions that would normally have been made under the domestic institutional arrangements of the country. We argue that this "contamination effect" generally prevents countries from containing violations of sovereignty caused by international agreements to narrow subsets of policy instruments. In fact, we establish as our second main result that within our benchmark model any international agreement that involves direct commitments over matters that are the internal affairs of a country must in general violate that country's sovereignty over at least as many policy instruments as it preserves.

Our first two results highlight an important distinction between international agreements that mitigate international externalities and international agreements that erode national sovereignty, and indicate that it can be possible to have the former without the latter and so allow national sovereignty and international efficiency to coexist in harmony. But questions remain as to (i) whether this harmony is likely to extend to environments beyond those captured by

our benchmark model, and (ii) the degree to which prominent channels of interdependence between countries can be identified that find representation in the environment described by our benchmark model. These questions are taken up in the second half of the chapter.

We consider a variety of settings that go beyond our benchmark model, and find that the harmony between national sovereignty and international efficiency does not always survive in these extended settings. For example, we identify a conflict between national sovereignty and international efficiency that arises whenever an externality variable is completely determined by the policy choices of a single country. This is because in this case the externality variable becomes this country's internal affairs according to our definition, and subjecting the externality variable to the constraints of an international agreement (which would be necessary to achieve international efficiency in this case) would therefore violate the country's sovereignty.

Another particularly salient extension of the benchmark model that we consider is to a world of "small" countries. When all countries are small in relation to the externality variable, we show that each country's contribution to the externality variable again becomes its internal affairs. This in turn implies that, when all countries are small, any international agreement that constrains countries from pursuing their Nash policy choices must violate their sovereignty. Accordingly, as we demonstrate, whether or not the harmony between national sovereignty and international efficiency described above survives in a world of small countries hinges on whether governments agree or disagree in the Nash equilibrium over the direction that they would like the externality variable to move. If all governments agree, then the Nash equilibrium in the small-country case is inefficient and an international agreement will be required to reach the efficiency frontier, implying necessarily that national sovereignty and international efficiency will stand in conflict in this case. However, if there is disagreement, then the Nash equilibrium in the small-country case is efficient, and in this case the harmony between national sovereignty and international efficiency identified above survives in a world of small countries. Of course, which of these two cases is applicable will depend on the nature of the externality variable under consideration, but as we later demonstrate, the latter case has special significance in the context of international trade agreements.

Besides being of interest in their own right, these extensions of our benchmark model highlight an important feature of our approach: rather than tailoring our definition of sovereignty on a case-by-case basis so that national sovereignty is necessarily in harmony with international efficiency in all circumstances, we propose a formal definition of sovereignty and then

evaluate the circumstances where a tradeoff between sovereignty and international efficiency is unavoidable.¹ As our benchmark model and its extensions reveal, the range of settings in which national sovereignty and international efficiency can coexist in harmony is non-trivial but also not exhaustive.

To explore whether prominent channels of interdependence between countries can be identified that find representation in the environment described by our benchmark model, we then consider the issue of national sovereignty within the particular context of international trade agreements, and we show that the trade setting fits well within the environment described by our benchmark model. To this end, we begin by briefly reviewing the two-country two-good competitive general equilibrium trade model adapted to allow for the possibility of both tariff and domestic regulatory policy choices as developed in Bagwell and Staiger (2001). We establish that this model is a special case of our benchmark model, in which the international externality variable is the terms of trade, and in which each country's contribution to the determination of the externality is the quantity of imports it demands at a given terms of trade, and so all of the results described above apply.

This leads to our third main result: when the results of our benchmark model are interpreted within the context of our trade model, they indicate that the fundamental principles underlying GATT/WTO market access agreements offer a way to achieve internationally efficient policies without sacrificing national sovereignty, and that attaining international efficiency is consistent with maintaining national sovereignty in this setting regardless of whether (all) countries are big or small. We also extend our analysis of trade agreements to a multilateral setting, and establish that agreement to abide by a nondiscrimination principle such as the GATT/WTO MFN rule does not violate a country's sovereignty. And finally, we identify a critical role for MFN if countries are to achieve internationally efficient policies without sacrificing national sovereignty when some (but not all) countries are small. In particular, we find that a non-discrimination rule can allow countries to sidestep the efficiency/sovereignty tradeoff that would otherwise exist in this extended setting, and we suggest that the MFN requirement is therefore "complementary" to preserving small-country sovereignty in the following sense: the sovereignty of small countries can be preserved under an internationally efficient agreement only if that agreement abides by the MFN requirement. More broadly, our results therefore suggest that a non-discrimination

¹As we discuss later (see notes 4 and 13), in his reconsideration of the concept of sovereignty Jackson (2003) takes the first approach, and proposes a notion of "sovereignty modern" that according to Jackson is capable of maintaining harmony between sovereignty and international efficiency in the context of modern sensibilities.

rule coupled with a market access agreement can facilitate the attainment of internationally efficient outcomes that do not compromise national sovereignty. In light of our findings, we discuss the basic harmony between the underlying GATT/WTO principles and the maintenance of national sovereignty, and we suggest that this harmony may be at risk as a result of changes that are occurring within the WTO.

The chapter proceeds as follows. Section 2 describes the two-country benchmark model. Section 3 develops our formal definition of sovereignty, characterizes the nature and degree of sovereignty in the Nash equilibrium, and relates this characterization to notions of sovereignty in the international political economy literature. Section 4 considers how national sovereignty is affected under international agreements that adopt alternative designs within the benchmark model, while section 5 considers the issue of sovereignty within a number of extensions of the benchmark model. Section 6 establishes that the benchmark model and all its results can be given a trade interpretation, and extends the modeling environment to a multilateral setting to consider the implications of a non-discrimination rule for our sovereignty results. Section 7 concludes. An Appendix contains proofs not included in the body of the chapter.

2. A Benchmark Model

In this section we describe a benchmark model of international interdependence that is general enough to allow interdependence to take a wide variety of forms. Our benchmark model has two countries (territories), referred to respectively as the home and foreign country, in which private agents (home and foreign citizens) reside. Each country has a government, and each government is endowed with a set of (tax and/or regulatory) policy instruments, represented by the $1 \times I$ vector \mathbf{i} for the home government and the $1 \times I^*$ vector \mathbf{i}^* for the foreign government, that are applied by each government to activities within its territory. The objectives of the home and foreign governments are represented by the respective functions $G(\mathbf{i}, \tilde{x}(\mathbf{i}, \mathbf{i}^*))$ and $G^*(\mathbf{i}^*, \tilde{x}(\mathbf{i}, \mathbf{i}^*))$, with the equilibrium level of the “externality” variable $\tilde{x}(\mathbf{i}, \mathbf{i}^*)$ entering into each government objective function and embodying the nature of the policy spillovers between the two countries. The ability to represent government objectives in this way reflects an essential assumption of our benchmark model, namely, that there exists a well-defined channel (e.g., the level of a price or the quantity of a pollutant) through which the effect of each government’s policy choices on the other government’s objectives (the externality) travels.

Aside from global concavity assumptions on the G and G^* functions to ensure that second-

order conditions are globally satisfied, the only additional structure we impose in the benchmark model is that $\tilde{x}(\mathbf{i}, \mathbf{i}^*)$ is a well-behaved function defined implicitly according to

$$\begin{aligned} f(g(\mathbf{i}, x), g^*(\mathbf{i}^*, x), x) &= 0, \\ \text{with } f_g g_{i_k} &\neq 0 \text{ for some } k \text{ and} \\ f_{g^*} g_{i_m}^* &\neq 0 \text{ for some } m, \end{aligned} \tag{2.1}$$

where here and throughout we use subscripts on a function to denote the partial derivative of that function with respect to the subscripted argument. In effect, g represents the home country's "contribution" to the determination of the equilibrium level of the externality variable \tilde{x} , a contribution that is assumed to be impacted by at least one policy instrument of the home government (i.e., $g_{i_k} \neq 0$ for some k); and this contribution is defined for a given level of the externality once the home-country policy instruments are determined. An analogous interpretation holds for g^* . The function f then aggregates the contributions of the home and foreign country to determine the equilibrium level of the externality \tilde{x} according to $f(\cdot) = 0$, under the assumption that f is impacted by changes in either country's contribution (i.e., $f_g \neq 0 \neq f_{g^*}$).

As we confirm in section 6 below, this structure is consistent with a setting in which the interdependence across countries is purely pecuniary and takes the form of international trade, with (2.1) then amounting to a market-clearing condition. But this structure is general enough as well to include many other forms of interdependence.

For example, x might represent the density of the fish population in a common fishery, with g representing the home catch when the home fleet operates in the policy environment \mathbf{i} and faces a fish population density x , and with g^* representing the foreign catch when the foreign fleet operates in the policy environment \mathbf{i}^* and faces a fish population density x . In this setting, it would be natural that $g_x > 0$ and $g_x^* > 0$. The equilibrium density of the fish population, given the policy environment faced by home and foreign fleets, $\tilde{x}(\mathbf{i}, \mathbf{i}^*)$, is then determined according to (2.1). Alternatively, x might represent the temperature of the globe, with g representing the home country's carbon output when the home industry operates in the policy environment \mathbf{i} and faces a global temperature x , and with g^* representing the carbon output of the foreign country when the foreign industry operates in the policy environment \mathbf{i}^* and faces a global temperature x . Here it could be that $g_x \gtrless 0$ and $g_x^* \gtrless 0$ depending on circumstances. The equilibrium temperature of the globe, given the policy environment faced by home and foreign

industries, $\tilde{x}(\mathbf{i}, \mathbf{i}^*)$, is then determined according to (2.1). Or, under the assumption that each government cares about infant mortality within its own borders but the home government also cares directly about infant mortality in the foreign country, x could represent the rate of infant mortality in the foreign country, with g representing the home country's crib exports to the foreign country when the home industry operates in the policy environment \mathbf{i} and the foreign infant mortality rate is x , and with g^* representing the foreign country's crib sales in the foreign market when the foreign industry operates in the policy environment \mathbf{i}^* and the foreign infant mortality rate is x . Here again it could be that $g_x \gtrless 0$ and $g_x^* \gtrless 0$ depending on circumstances. The equilibrium infant mortality rate in the foreign country, given the policy environment faced by home and foreign industries, $\tilde{x}(\mathbf{i}, \mathbf{i}^*)$, is then determined according to (2.1).

Importantly, the structure in (2.1) rules out the possibility that $f_g g_{i_k} = 0$ for all k and/or that $f_{g^*} g_{i_m}^* = 0$ for all m , and thereby excludes cases where the externality variable x cannot be impacted by the policy instruments of some government. We will extend the benchmark model to consider such cases in a later section.

In the absence of any international agreements, we assume that each government makes choices over its own set of policy instruments, taking as given the policy levels of the other government. This problem defines, for each government, its best-response policy choice problem, and the solution to this problem defines its best-response policy choices. Specifically, the home government chooses its best-response policies by solving $\max_{\mathbf{i}} G(\mathbf{i}, \tilde{x}(\mathbf{i}, \mathbf{i}^*))$ taking \mathbf{i}^* as given, which using (2.1) we write in the equivalent form

$$\mathbf{Program\ 1:} \quad \text{For a given } \mathbf{i}^*, \quad \max_{\mathbf{i}, x} G(\mathbf{i}, x) \quad \text{s.t.} \quad f(g(\mathbf{i}, x), g^*(\mathbf{i}^*, x), x) = 0,$$

at the same time that the foreign government chooses its best-response policies by solving $\max_{\mathbf{i}^*} G^*(\mathbf{i}^*, \tilde{x}(\mathbf{i}, \mathbf{i}^*))$ taking \mathbf{i} as given, which using (2.1) we write in the equivalent form

$$\mathbf{Program\ 1^*}: \quad \text{For a given } \mathbf{i}, \quad \max_{\mathbf{i}^*, x} G^*(\mathbf{i}^*, x) \quad \text{s.t.} \quad f(g(\mathbf{i}, x), g^*(\mathbf{i}^*, x), x) = 0.$$

Substituting the constraint into the objective function for each program, the I first-order conditions associated with Program 1 that define the home government's best-response policy choices are given by

$$G_{i_k} + G_{\tilde{x}} \tilde{x}_{i_k} = 0 \quad \text{for } k = 1, \dots, I, \tag{2.2}$$

while the I^* first-order conditions associated with Program 1* that define the foreign government's best-response policy choices are given by

$$G_{i_k}^* + G_{\tilde{x}}^* \tilde{x}_{i_k}^* = 0 \quad \text{for } k = 1, \dots, I^*. \tag{2.3}$$

The joint solutions to (2.2) and (2.3) define the Nash equilibrium of the benchmark model, which throughout we assume exists and is unique.

We next characterize the international efficiency frontier and evaluate the efficiency properties of the Nash equilibrium. We define the international efficiency frontier with respect to the objectives of each government. Accordingly, the international efficiency frontier solves the following program:

$$\begin{aligned} \mathbf{Program\ 2} : \max_{\mathbf{i}, \mathbf{i}^*, x} G(\mathbf{i}, x) \\ \text{s.t. (i) } G^*(\mathbf{i}^*, x) \geq \bar{G}^*, \text{ and} \\ \text{(ii) } f(g(\mathbf{i}, x), g^*(\mathbf{i}^*, x), x) = 0. \end{aligned}$$

Using Program 2, and letting $k = 1$ denote a domestic and foreign policy instrument for which $g_{i_k} \neq 0$ and $g_{i_k}^* \neq 0$, it is direct to derive the first-order conditions that characterize the international efficiency frontier:

$$G_{i_k} = \frac{g_{i_k} G_{i_1}}{g_{i_1}} \quad \text{for } k = 2, \dots, I, \quad (2.4)$$

$$G_{i_k}^* = \frac{g_{i_k}^* G_{i_1}^*}{g_{i_1}^*} \quad \text{for } k = 2, \dots, I^*, \text{ and} \quad (2.5)$$

$$G_{i_1} G_{i_1}^* + G_{i_1} G_x^* \tilde{x}_{i_1}^* + G_{i_1}^* G_x \tilde{x}_{i_1} = 0, \quad (2.6)$$

along with the complementary slackness conditions ensuring that the Kuhn-Tucker multiplier on constraint (i) of Program 2 is non-negative:

$$\begin{aligned} -\frac{G_x}{G_x^*} + \frac{G_{i_1} [f_g g_x + f_{g^*} g_x^* + f_x]}{G_x^* f_g g_{i_1}} &\geq 0 \text{ if } G_x^* \neq 0 \text{ and} \\ \frac{G_{i_1} f_g g_{i_k}^*}{G_{i_k}^* f_g g_{i_1}} &\geq 0 \text{ for each } k \text{ for which } G_{i_k}^* \neq 0. \end{aligned} \quad (2.7)$$

The efficiency properties of the Nash equilibrium may now be assessed. Using (2.2) and (2.3), whose joint solutions define the Nash equilibrium in our benchmark model, we may state:

Proposition 1. *The Nash equilibrium of the benchmark model is inefficient if and only if $G_x \neq 0$ and $G_x^* \neq 0$ at the Nash policy choices.*

Proof: See Appendix.²

According to Proposition 1, the Nash equilibrium is inefficient – and hence a potential role for international agreements arises in our benchmark model – if and only if $G_x \neq 0$ and $G_x^* \neq 0$ at the Nash policy choices. We will refer to the case where $G_x \neq 0$ and $G_x^* \neq 0$ at the Nash policy choices as the case where the home and foreign countries are *mutually interdependent*. As our central purpose is to consider the implications of voluntary (i.e., mutually beneficial) international agreements for national sovereignty, and as the case of mutually interdependent countries is the only case in which voluntary international agreements can arise, this case will be a primary focus of the analysis to follow.

3. What is Sovereignty?

With the essential elements of our benchmark model described, we now turn to develop a definition of sovereignty within the context of this model. Defining sovereignty is not a simple task. On the one hand, to be operational, our definition of sovereignty must be amenable to formal analysis. On the other hand, to be relevant, our definition of sovereignty must capture elements that feature prominently in the common usage of this term. This latter requirement is particularly difficult, because the international political economy literature within which sovereignty has been most discussed is not always clear about the precise meaning of the term and, when clear, does not always adopt a uniform meaning. In fact, Krasner (1999) identifies four distinct ways in which the term “sovereignty” has been commonly used in this literature. Krasner refers to these as *domestic sovereignty*, *international legal sovereignty*, *interdependence sovereignty*, and *Westphalian sovereignty*. Domestic sovereignty refers to the organization and effectiveness of political authority within the state. International legal sovereignty refers to the mutual recognition of states. Interdependence sovereignty refers to the scope of activities over which states can effectively exercise control. And Westphalian sovereignty reflects as its central premise the rule of nonintervention in the internal affairs of other states.

Our starting point for defining sovereignty is the Westphalian norm of “non-intervention in the internal affairs of other states” (Krasner, 1999, p. 20). To formalize this definition, we

²The required conditions for inefficiency of the Nash equilibrium in the benchmark model would be weakened if the ability to make explicit international transfers were introduced. This can be done without changing the nature of any of our results, but we prefer to keep explicit international transfer instruments out of our model for simplicity. We also consider several extensions of the benchmark model in a later section where the Nash inefficiencies take more complicated forms.

must define “non-intervention” and “internal affairs.”

What are the internal affairs of a state? One answer might be to equate internal affairs directly with those matters that no one outside the state cares about. But as a basis for defining national sovereignty, this approach has limitations, because it leads to a notion of sovereignty that is defined directly by the preferences of external actors: if an external actor decides that it cares about an issue over which a state is making a choice, then that issue is by this definition no longer an internal affair of the state.³ We therefore look to the international political economy literature for guidance. In this literature, the internal affairs of a state are synonymous with its “domestic authority structures,” a phrase that in turn has been interpreted to mean the state’s authority to determine the institutions and policies that apply within its territorial boundaries. For example, according to Krasner (1999, p. 20), the concept of Westphalian sovereignty can be characterized as

“...an institutional arrangement for organizing political life that is based on two principles: territoriality and the exclusion of external actors from domestic authority structures. Rulers may be constrained, sometimes severely, by the external environment, but they are still free to choose the institutions and policies they regard as optimal. Westphalian sovereignty is violated when external actors influence or determine domestic authority structures.”

Here Krasner is drawing a distinction between the effectiveness with which control over outcomes can be exerted on the one hand, and the authority to choose institutions and policies on the other. The former is a concern of domestic sovereignty and interdependence sovereignty, but it is the latter that defines the internal affairs with which Westphalian sovereignty is concerned. In the international political economy literature, then, a nation’s internal affairs – over which, according to Westphalian sovereignty, it must enjoy freedom from intervention by external actors – are considered to be its choice of domestic institutions and the operation of these institutions to translate the preferences of its citizens into policy choices.

To draw a connection between this notion of internal affairs and our benchmark model, we observe that our home and foreign government objectives G and G^* respectively can be said to reflect the particular set of “domestic authority structures” relevant for determining the levels

³For a related discussion of the practical difficulties associated with the implementation of Mill’s “harm principle,” see, for example, Gray (1991).

of policy instruments within the territory of that government. For example, domestic authority may be concentrated in the hands of one individual, whose preferences are then the government objective function for that country. If that individual is subjected to lobbying by interest groups, then these interest groups also comprise a part of the domestic authority structure, and the government objective function for that country will reflect as well the influence wielded by these interest groups. Alternatively, domestic authority may be dispersed in the hands of the electorate and take the form of a direct democracy, in which case under appropriate assumptions the preferences of the median voter are then the government objective function for that country. Or domestic authority over different policies may be allocated across different domestic institutions: as long as coordination across domestic institutions (e.g. bargaining among them) is possible, our representation of government objectives allows a valid description of the domestic policy environment in this setting as well. The point is, each government's objective function as we have defined it reflects both the underlying preferences of the citizens of that country and the domestic authority structures under which those preferences are translated into choices over policy instruments in that country. According to the meaning of Westphalian sovereignty in the international political economy literature, then, within the context of our model the internal affairs of the home (foreign) country are embodied in its choice of G (G^*).

As for what constitutes intervention by external actors, Krasner (1999, Chapters 6 and 7) observes that coercion (as in international armed conflict) has frequently resulted in constitutional changes that explicitly alter the domestic institutions of a country and thereby violate its Westphalian sovereignty. Such explicit changes in domestic institutions could be interpreted within our benchmark model as alterations in the G and/or G^* functions, reflecting for example the forced removal of a dictator and the introduction of democratic institutions. However, our focus here is not on armed conflict, but rather on voluntary international agreements. Hence, violations of Westphalian sovereignty that would result in changes in the G and/or G^* functions are not our central concern.

But Krasner (1999, p. 22) observes that invitation (as in international contracts and conventions) can also violate Westphalian sovereignty, not necessarily by explicitly altering domestic institutions, but by "...subjecting internal authority structures to external constraints." Rabkin (1998, p. 34) puts the point slightly differently: (Westphalian) sovereignty is violated by international commitments that "distort or derange the normal workings of our own system...". In effect, international commitments need not alter the domestic institutions of a country in order

to violate its Westphalian sovereignty: international commitments that distort the operation of domestic institutions will also violate Westphalian sovereignty.

Implicit in the above discussion is the notion that specific commitments arising out of voluntary international agreements would not ordinarily be viewed as violations of Westphalian sovereignty. For Westphalian sovereignty to be violated by invitation, a “deeper” intervention into the internal affairs of the state is required. Even here though, both Krasner (1999) and Rabkin (1998) suggest that there are limits to the appropriate matters for international agreement, and that a nation’s Westphalian sovereignty would be violated by negotiated commitments over policies that cross these limits and stray into “sufficiently domestic” affairs. For instance, Krasner (1999, pp. 146-148) observes that the IMF routinely violates the norm of Westphalian sovereignty, in part because “...A country entering into negotiations with the IMF could basically consider any aspect of its domestic economic policy open to discussion.” Similarly, in the Preface to his book, Rabkin (1998, p. x) states that efforts to delineate the appropriate limits of international commitments are “...particularly urgent now because, in the absence of any clear understandings on the matter, we seem to be letting international agreements and international authorities determine more and more of our policies.” Neither Krasner nor Rabkin offer a precise method for defining the limits of proper subjects of international negotiation, though Rabkin (pp. 69-70) proposes several criteria.

Summarizing, three key features of Westphalian sovereignty that seem especially relevant in the context of voluntary international agreements can be identified from our discussion of the international political economy literature to this point: first, commitments that result from voluntary international agreements do not necessarily violate Westphalian sovereignty; second, international commitments over policies that concern “sufficiently domestic” affairs (i.e., internal affairs) *do* violate Westphalian sovereignty; and third, international commitments that distort the normal operation of domestic institutions also violate Westphalian sovereignty. We wish to construct a definition of sovereignty that reflects these three features.

To accomplish this, we maintain as our essential focus the Westphalian norm of non-intervention in the internal affairs of other states. And below we adopt a definition of non-intervention that is well-reflected in the discussion above. But in proposing a formal definition of internal affairs, we augment the Westphalian emphasis on authority over the determination of institutions and policies, and add to this an emphasis on authority and control over the determination of outcomes and therefore payoffs as well, all evaluated from the perspective

of the Nash policy equilibrium of our model. In effect, our definition of sovereignty combines elements of authority with elements of control/effectiveness, and in so doing delivers a notion of sovereignty that exhibits traditional features (Krasner, 1999, p. 10) of Westphalian sovereignty (authority over institutions and policies), interdependence sovereignty (effective control over cross-border activities) and domestic sovereignty (authority and effective control over activities within the territory). As a result, the characterization of a country's internal affairs according to our definition will depend on the nature of interdependence across countries. With internal affairs so defined, our broad approach is then to characterize the normal operation of a country's domestic institutions in the domain of its internal affairs, and to say that a violation of sovereignty occurs whenever an international agreement leads the government of a country to make external commitments over matters that (i) concern the country's internal affairs or (ii) alter (and therefore influence/distort/derange) the normal operations of the country's domestic institutions within the domain of its internal affairs.

Our approach to defining national sovereignty has both advantages and disadvantages. On the minus side, our approach does not conform precisely to any one of the four notions of sovereignty commonly discussed in the international political economy literature: as we noted above, it combines elements of a number of these notions. On the plus side, however, our approach admits several advantages. First and foremost, as we demonstrate below this approach provides an analytically tractable way to capture the three key features of Westphalian sovereignty identified above. Second, because a country's internal affairs according to our definition will depend on the nature of interdependence across countries, as the nature of interdependence changes so too will the scope of a country's internal affairs and hence the domain of its sovereign rights.⁴ And finally, by augmenting the Westphalian focus on authority over

⁴In arguing for the need to update the traditional Westphalian concept of sovereignty, Jackson (2003) states:

... Much has been said and written about "globalization"; despite being an ambiguous term of controversial connotation, it is reasonably well understood to apply to the exogenous world circumstances of economic and other forces that have developed in recent decades owing, in major part, to the sharply reduced costs and time required for the transport of goods (and services), and similar reductions in costs and time requirements for communication. These circumstances have led to new structures of production; they, in turn, have resulted in greatly enhanced (and sometimes dangerous) interdependence, which we can do little to remedy and which often renders the older concepts of "sovereignty" or "independence" fictional. ... these circumstances often demand action that no single nation-state can satisfactorily carry out, and thus require some type of institutional "coordination" mechanism. In some of these circumstances, therefore, a powerful tension is generated between traditional core "sovereignty," on the one hand, and the international institution, on the other hand. (Jackson, 2003, p 784, footnotes omitted)

institutions/policies with a focus as well on authority over outcomes/payoffs, our approach may facilitate a more direct link to issues of *accountability* in an interdependent world than do any of the existing notions of sovereignty taken separately, and may thus be of some interest in its own right. For example, a government might maintain authority over its institutions and policies (and therefore maintain Westphalian sovereignty) and yet claim that it cannot be held accountable for its choices, as a consequence of a “race to the bottom” that external constraints have forced upon it. But in matters where the government maintains control over outcomes/payoffs as well (and hence in matters that are also the country’s internal affairs according to our definition), this possibility of avoiding accountability by appealing to external constraints does not arise. As a consequence, our approach to defining national sovereignty can be used to forge a tighter link between international agreements that can be said to avoid an erosion of sovereignty and those that can be said to avoid an erosion of accountability.

3.1. Sovereignty Defined

We now proceed to develop our formal definitions in detail. To this end, we return to the benchmark model presented in section 2, and consider the Nash policy choices made by each government. It is with respect to Nash choices, unconstrained by any international commitments, that a nation’s internal affairs can be defined. As noted above, we adopt a definition of internal affairs that equates the internal affairs of a country with (i) the choice of domestic authority structures under which the preferences of its citizens are translated into choices over policy instruments in that country, as embodied in the home and foreign government objective functions G and G^* respectively, and (ii) the matters in which its government has control or “sole authority” over outcomes/payoffs (in the Nash equilibrium).

To develop this definition, we consider alternative representations of each government’s best-response policy choice problem (that is, alternative representations of Program 1 and Program 1* in section 2) that *partition* this problem into a sequence of sub-problems, each of which can be considered a choice problem of its own, and where the solution to any such representation

Jackson’s approach is to propose an updated concept of Westphalian sovereignty that he terms “sovereignty-modern,” and which is meant to be more consistent with international efficiency and the need for international policy coordination in the modern world. Our approach provides a formal definition of sovereignty which achieves some of what Jackson has in mind, in that according to our definition the domain of sovereignty will evolve as the nature of international interdependence evolves; but we do not tailor our definition of sovereignty on a case-by-case basis to necessarily be in harmony with international efficiency, and instead evaluate formally the circumstances when a tradeoff between maintaining national sovereignty and achieving international efficiency can be avoided versus when this tradeoff will necessarily arise.

yields the original best-response policy choices of the government characterized by (2.2) and (2.3). Our approach is to use these partitions to identify the maximal set of choice problems over which a government can be said to exercise sole authority in the Nash equilibrium, and thereby (together with its choice of government objective function) identify the country’s internal affairs according to our definition.

We begin by defining a partition of a government’s best-response policy choice problem:

Definition 1. A *partition* \mathcal{P} of a government’s best-response policy choice problem is any sequence of choice problems whose solution yields the best-response policy choices of the government.

According to its definition, each element of \mathcal{P} is a choice problem for the government, and when the government has solved each of these choice problems it arrives at the original best-response policies.⁵ We also need a definition of authority:

Definition 2. A government has *sole authority* in a choice problem if and only if its payoff in that choice problem is independent of the actions of “external actors.”

By the actions of external actors, we mean the setting of all policy instruments by the government of the other country and all decisions by private agents in the other country. Accordingly, a government has authority in a choice problem provided that its policy choices, together with the decisions of private agents operating within its borders, fully determine its payoffs in that choice problem.

We next define internal affairs and external affairs, conditional on the partition \mathcal{P} under consideration:

Definition 3. For any partition \mathcal{P} of its government’s best-response policy choice problem, a country’s *\mathcal{P} -internal affairs* consist of (i) the choice of domestic authority structures under which the preferences of its citizens are translated into choices over policy instruments in that country, as embodied in its government objective function, and (ii) the collection of choice problems in \mathcal{P} over which the government has sole authority. The country’s *\mathcal{P} -external affairs* are the remaining choice problems in \mathcal{P} .

⁵Notice that, as the home government’s best-response policy choices are defined by the vector of domestic policies \mathbf{i} that solve 2.2, the choice problems in any partition \mathcal{P} of the home government’s best-response policy choice problem cannot include (trivial or otherwise) choices over foreign policies in \mathbf{i}^* , and similarly the choice problems in any partition \mathcal{P} of the foreign government’s best-response policy choice problem cannot include (trivial or otherwise) choices over home policies in \mathbf{i} .

In other words, conditional on a partition \mathcal{P} of its government's best-response policy choice problem, a country's \mathcal{P} -internal affairs are its choice of government objective function and the choice problems over which its policy choices fully determine its payoffs, and its \mathcal{P} -external affairs are the remaining choice problems in the partition \mathcal{P} .

Definition 3 permits any choice problem to be assigned either to a country's internal affairs or to its external affairs. But these choice problems are defined relative to a specific partition of the government's best-response policy choice problem, and there are many such partitions. Hence, as a general matter, the internal affairs of a country as we have defined them in definition 3 will depend on the partition under consideration, a dependence we have indicated with the terms \mathcal{P} -internal affairs and \mathcal{P} -external affairs.

However, if there exists a way to partition the government's best-response policy choice problem so that a *minimal* set of external affairs can be identified (in the sense that, as we formalize below, there is no partition under which a country's external affairs would be a proper subset of this set), then we prefer to select this partition among all possible partitions for purposes of defining a county's internal and external affairs. The minimal partition, when it exists, seems the reasonable partition on which to focus, because it would identify the smallest collection of choice problems over which the government does not enjoy sole authority in the Nash equilibrium, and by implication as well the largest collection of choice problems over which the government does enjoy sole authority in the Nash equilibrium. And so in this case, the reference to dependence on the particular partition under consideration can be suppressed, and we can refer simply to internal and external affairs.

We next develop this preferred definition of internal and external affairs. To state this definition, we say that a collection of choice problems \hat{s} contained in the partition $\hat{\mathcal{P}}$ is a *subset* of the collection of choice problems \tilde{s} contained in the partition $\tilde{\mathcal{P}}$ provided that there exists a collection of choice problems \tilde{s}' such that $\{\{\tilde{\mathcal{P}} \setminus \tilde{s}\} \cup \tilde{s}'\}$ is also a partition of the government's best-response policy choice problem and every choice problem in \hat{s} is also in \tilde{s}' . Intuitively, \hat{s} is a subset of \tilde{s} if there is a way to write \tilde{s} – which we denote \tilde{s}' , and with $\{\{\tilde{\mathcal{P}} \setminus \tilde{s}\} \cup \tilde{s}'\}$ also a partition of the government's best-response policy choice problem – so that \hat{s} and \tilde{s}' can be compared on a choice-problem by choice-problem basis, and when this comparison is made every element of \hat{s} (every choice problem in \hat{s}) is also in \tilde{s}' ; this would mean that \tilde{s} can be thought of as \hat{s} plus “stuff.”⁶ And we say that a collection of choice problems \hat{s} contained in

⁶Note that according to our definition, it is possible that \hat{s} is a subset of \tilde{s} and yet the cardinality of \hat{s} is

the partition $\hat{\mathcal{P}}$ is *equivalent* to the collection of choice problems \tilde{s} contained in the partition $\tilde{\mathcal{P}}$ provided that there exists a collection of choice problems \tilde{s}' such that $\{\{\tilde{\mathcal{P}} \setminus \tilde{s}\} \cup \tilde{s}'\}$ is also a partition of the government's best-response policy choice problem and a choice problem is in \hat{s} if and only if it is also in \tilde{s}' .

We may now define a minimal partition:

Definition 4. A *minimal partition* $\hat{\mathcal{P}}$ of a government's best-response policy choice problem is a partition for which the government's $\hat{\mathcal{P}}$ -external affairs are a subset of its \mathcal{P} -external affairs for all \mathcal{P} .

And with this we arrive at our preferred definition of internal and external affairs:

Definition 5. If there exists a minimal partition $\hat{\mathcal{P}}$ of its government's best-response policy choice problem, then a country's *internal affairs* consist of (i) the choice of domestic authority structures under which the preferences of its citizens are translated into choices over policy instruments in that country, as embodied in its government objective function, and (ii) its $\hat{\mathcal{P}}$ -internal affairs; and the country's *external affairs* are its $\hat{\mathcal{P}}$ -external affairs.

Notice that it is possible that a set of partitions may qualify as minimal partitions, but each partition in this set must (i) be payoff equivalent, because all partitions must lead to the same best-response policy choices, and (ii) deliver the same characterization of internal and external affairs according to definitions 4 and 5. Hence, this possible non-uniqueness is immaterial for our purposes.⁷

To make use of definition 5, we must establish the existence of minimal partitions of the home and foreign government best-response choice problems in our benchmark model. To this

greater than the cardinality of \tilde{s} . This possibility would simply reflect that \hat{s} had not been written in a form that had allowed direct comparison with \tilde{s} on a choice-problem by choice-problem basis (and that by such rewriting it became possible to show that in fact \hat{s} was a subset of \tilde{s} according to our definition)

⁷As a general matter, we do not have an existence proof for a minimal partition, which is why we state the definition of internal and external affairs conditional on the existence of a minimal partition. But as we next show, we are able to establish existence in the formal setting of our benchmark model.

end, consider the alternative 2-step representation of Program 1:

$$\begin{aligned}
 \mathbf{Program\ 1}' : \quad & \text{Step 1. For a given } (g, x) : \quad \max_{\mathbf{i}} G(\mathbf{i}, x) \\
 & \text{s.t. } [g(\mathbf{i}, x) - g] = 0. \\
 & \text{Step 2. For a given } \mathbf{i}^* : \max_{g, x} L(\hat{\mathbf{i}}(g, x), g, x) \\
 & \text{s.t. } f(g, g^*(\mathbf{i}^*, x), x) = 0,
 \end{aligned}$$

where $\hat{\mathbf{i}}(g, x)$ is the solution from Step 1 and L is the Step-1 Lagrangean.⁸ The Step-1 choice problem in Program 1' is solved conditional on a given level of the “externality” variable x and the home-country’s “contribution” g to the externality variable, and has the home government making its preferred choices over domestic policy instruments \mathbf{i} so as to deliver this contribution. The Step-2 choice problem has the home government then making its preferred choices over g and x subject to the constraint placed on its choices which is implied by a vector of foreign policy instruments \mathbf{i}^* .

Our first result is that Program 1' is indeed an equivalent way of characterizing the home-government’s best-response policies in Program 1. We record this in:

Lemma 1. Program 1' is a partition of Program 1.

Proof: See Appendix.

We prove Lemma 1 by establishing that the first-order conditions associated with Program 1' are given by (2.2), the first-order conditions associated with Program 1, hence the solution of Program 1' yields the best-response policy choices of the home government. While we have developed this partition from the perspective of the home government’s problem Program 1, an exactly analogous partition (which we denote henceforth by Program 1*') can be developed for the foreign government’s problem Program 1*. For future reference, we denote by \mathcal{P}_0 the partition of the home-government’s best-response policy choice problem defined by Program 1', and by \mathcal{P}_0^* the partition of the foreign-government’s best-response policy choice problem defined by Program 1*'.⁸

We next use Program 1' (and the analogous Program 1*' of the foreign government) to assess the internal affairs of each country according to the partitions \mathcal{P}_0 and \mathcal{P}_0^* . It is immediate

⁸It should be understood that the set of g and x for which Step-1 is solved are those for which there exists a vector of instruments \mathbf{i} such that $[g(\mathbf{i}, x) - g] = 0$ is satisfied.

from Program 1' that the home government has sole authority over its payoff in the choice problem defined by Step 1, and hence the Step-1 choice problem concerns the home-country's \mathcal{P}_0 -internal affairs for this alternative partition. We refer henceforth to the home-government's Step- 1 choice problem in the partition \mathcal{P}_0 as its *choices over $\mathbf{i}(g, x)$* , and as noted above we denote its Step-1 choices by $\hat{\mathbf{i}}(g, x)$. On the other hand, the home government would have sole authority over its payoff in the choice problem defined by Step 2 – and hence the Step-2 choice problem would concern the home-country's \mathcal{P}_0 -internal affairs for this alternative partition – if and only if the home government faced a non-binding constraint in its Step-2 choice problem, which cannot occur as long as G_x is non-zero.⁹ We refer henceforth to the home-government's Step-2 choice problem in the partition \mathcal{P}_0 as its *choices over g and x* . Completely analogous statements hold for the foreign government. Hence we have:

Lemma 2. The home country's \mathcal{P}_0 -internal affairs consist of its choice of G and its government's choices over $\mathbf{i}(g, x)$, and also its government's choices over g and x if and only if $G_x = 0$ when evaluated at the Nash policy choices. The foreign country's \mathcal{P}_0^* -internal affairs consist of its choice of G^* and its government's choices over $\mathbf{i}^*(g^*, x)$, and also its government's choices over g^* and x if and only if $G_x^* = 0$ when evaluated at the Nash policy choices.

Finally, we now establish that the partitions \mathcal{P}_0 and \mathcal{P}_0^* are minimal partitions of the home and foreign government best-response choice problems, respectively. This is stated in:

Lemma 3. The partition \mathcal{P}_0 is a minimal partition of the home-government's best-response choice problem, and the partition \mathcal{P}_0^* is a minimal partition of the foreign-government's best-response choice problem.

Proof: If $G_x = 0$ and $G_x^* = 0$ when evaluated at the Nash policy choices, then according to \mathcal{P}_0 and \mathcal{P}_0^* there are no choices for either the home- or the foreign- government that concern its external affairs, and so \mathcal{P}_0 and \mathcal{P}_0^* must be minimal partitions in this case. Consider, then, the case in which $G_x \neq 0$ when evaluated at the Nash policy choices. According to \mathcal{P}_0 , the home government's choices over g and x are its external affairs in this case. Suppose that \mathcal{P}_0 is not a minimal partition. Then there must exist a partition \mathcal{P}' of the home government's

⁹Forming the Lagrangean associated with the home government's Step-2 problem, it is direct to confirm that the expression for the Lagrange multiplier is $\lambda = G_x/[f_g g_x + f_{g^*} g_x^* + f_x]$, from which the statement above can be confirmed.

best-response choice problem in which the home government's choices over g and x are not both included in its external affairs. In the partition \mathcal{P}' , it must then be possible for the home government to alter either g or x or both g and x in at least one choice problem contained in its internal affairs. Consider, then, any choice problem contained in the home government's \mathcal{P}' -internal affairs for which either g or x or both g and x can be altered. With \mathbf{i}^* fixed and with $f_g \neq 0$ by assumption, g and x cannot be determined independently. Therefore, in this choice problem, the home government must face a constraint of the form $f(g, g^*(\mathbf{i}^*, x), x) = 0$ if it chooses g directly, or a constraint of the form $f(g(\mathbf{i}, x), g^*(\mathbf{i}^*, x), x) = 0$ if it instead chooses g indirectly through its choice of elements of \mathbf{i} . But either way, the constraint ensures that the home government does not have sole authority in this choice problem – and this is inconsistent with the claim that this choice problem concerns a matter of internal affairs for the home government – as long as $G_x \neq 0$ when evaluated at the Nash policy choices, which is the case under consideration. Hence, we have derived a contradiction, and so \mathcal{P}_0 must be a minimal partition. An analogous argument holds for the foreign government. QED

As a result of Lemmas 1-3, we may then state:

Proposition 2. If countries are mutually interdependent so that $G_x \neq 0$ and $G_x^* \neq 0$ at the Nash policy choices, then in the benchmark model the home country's internal affairs are its choice of G and its government's choices over $\mathbf{i}(g, x)$, and the foreign country's internal affairs are its choice of G^* and its government's choices over $\mathbf{i}^*(g^*, x)$; and choices over g and x (g^* and x) represent the external affairs of the home- (foreign-) country. If countries are not mutually interdependent, then in the benchmark model: (a) if $G_x = 0$ at the Nash policy choices, the home country's internal affairs include as well its government's choices over g and x ; and (b) if $G_x^* = 0$ at the Nash policy choices, the foreign country's internal affairs include as well its government's choices over g^* and x .

According to Proposition 2, when countries are mutually interdependent, the matters that concern the internal affairs of each country are the domestic authority structures under which the preferences of its citizens are translated into choices over policy instruments in that country, as embodied in its government objective function, and its government's choices among the set of policy combinations that are consistent with a given contribution to and level of the externality variable, since its payoff in this choice problem is independent of the actions of external actors. Notice that the policy choices made by each government in matters that concern its country's

internal affairs, namely $\hat{v}(g, x)$ for the home government and $\hat{v}^*(g^*, x)$ for the foreign government, reflect both the underlying preferences of the citizens of that country and the normal operation of that country's domestic institutions under which those preferences are translated into choices over policy instruments. The external affairs of each country then consist of its government's choices over the country's contribution to the determination of the externality variable and the equilibrium level of the externality variable, since its payoff in this choice problem depends on the actions of external actors. And when countries are not mutually interdependent, *all* of their choices may become their internal affairs (and will be, when both $G_x = 0$ and $G_x^* = 0$) according to Proposition 2. In light of Proposition 1, an implication of Proposition 2 is that countries operating in an environment of mutual interdependency where a potential role for international agreements exists can claim national sovereignty only over a diminished range of choices (their internal affairs) even when they are not part of any international agreement.

In light of Proposition 2, we are now ready to define external intervention as it relates to voluntary international agreements:

Definition 6. An international agreement *subjects the internal affairs of a country to external constraints* if and only if: (i) the government of that country makes commitments in the agreement over matters that concern its internal affairs; and/or (ii) the agreement has the effect of altering the choices in any choice problem that concerns the country's internal affairs.

With this definition, we then say that a country's *sovereignty is violated by an international agreement* when its internal affairs are subjected to external constraints by that agreement.¹⁰

Returning now to the three key features that we described in section 3 regarding Westphalian sovereignty as it relates to voluntary international agreements, we observe that our formalization of national sovereignty reflects each of these features. First, commitments that result from voluntary international agreements do not necessarily violate sovereignty since, according to Proposition 1, international agreements over g , x and g^* do not violate sovereignty as long as governments are mutually interdependent. Second, international commitments over policies that concern "sufficiently domestic" affairs (i.e., internal affairs) *do* violate Westphalian sovereignty since, according to Proposition 2, negotiated commitments over the elements of $\hat{v}(g, x)$ and/or $\hat{v}^*(g^*, x)$ always violate sovereignty. And third, international commitments that distort

¹⁰A distinction may be made between explicit and tacit agreements. We focus here on how sovereignty might be violated by an explicit agreement. A related possibility is that governments may achieve a tacit agreement that emerges from repeated interaction, for example. We do not explore that possibility here.

the normal operation of domestic institutions also violate sovereignty, if these distortions result in unilateral policy choices that do not conform to the corresponding elements of $\hat{i}(g, x)$ and/or $\hat{i}^*(g^*, x)$.

4. Sovereignty, International Agreements and Efficiency

We now make use of our definition of sovereignty. In this section we consider international agreements of various kinds to explore within our benchmark model the nature of the tradeoff between the preservation of sovereignty and the attainment of international efficiency. Throughout we assume that, subsequent to the conclusion of negotiations of an international agreement, each government chooses its best-response policies unilaterally given the policies of the other government and subject to the constraints placed on it by the international agreement.

We first consider the possibility that international efficiency might be attained without violating national sovereignty. When countries are not mutually interdependent, it is direct from Propositions 1 and 2 that this is possible, as long as countries avoid being constrained by any international commitments. This follows because when countries are not mutually interdependent the Nash equilibrium is efficient by Proposition 1, and by Proposition 2 all of at least one country's choices are its internal affairs. When countries are mutually interdependent, however, an international agreement is required to achieve efficiency, as Proposition 1 implies. The question is whether an international agreement can in this case be designed to achieve international efficiency while not violating national sovereignty. The answer to this question within our benchmark model is affirmative, as we record in the next proposition:

Proposition 3. In the benchmark model, there is no inherent conflict between international efficiency and national sovereignty. When countries are not mutually interdependent, they can maintain their national sovereignty and attain international efficiency by avoiding international commitments. When countries are mutually interdependent, they can maintain their national sovereignty and attain international efficiency by negotiating commitments over g , x and g^* .

Proof: We focus on the case of mutually interdependent countries. If the home and foreign government negotiate commitments over g , x and g^* , then subsequent to the conclusion of negotiations the home government will, for the negotiated level of (g, x) , choose its vector of policies \mathbf{i} subject to the implied constraint $[g(\mathbf{i}, x) - g] = 0$, and therefore solve Step 1 of Program 1', and similarly the foreign government will, for the negotiated level of (g^*, x) , choose its vector

of policies \mathbf{i}^* subject to the implied constraint $[g^*(\mathbf{i}^*, x) - g^*] = 0$, and therefore solve Step 1 of Program 1 * . With L and L^* the Step-1 Lagrangeans for the home and foreign government, respectively, negotiations over g , x and g^* then solve the following program for some \bar{L}^* :

$$\begin{aligned} \mathbf{Program\ 3} : \max_{g, x, g^*} & L(\mathbf{i}(g, x), g, x) \\ \text{s.t. (i)} & L^*(\mathbf{i}^*(g^*, x), g^*, x) \geq \bar{L}^*, \text{ and} \\ & \text{(ii) } f(g, g^*, x) = 0. \end{aligned}$$

It is now direct to derive that the first-order and complementary slackness conditions associated with Program 3 are identical to those associated with Program 2, and given by (2.4)-(2.6) and (2.7) respectively. Hence, negotiations over g , x and g^* achieve the international efficiency frontier without violating national sovereignty. **QED**

Proposition 3 implies that, in principle, countries need not confront a choice between preserving their sovereignty and attaining international efficiency when they negotiate international agreements, as long as they (a) negotiate agreements that impose commitments only on countries with which they are mutually interdependent, and (b) limit negotiations to direct commitments over the level of the externality (x) and each country's contribution to the externality (g and g^*). Intuitively, the international externality is the source of both the mutual interdependence across countries and the inefficiency associated with their unilateral policy choices, and it also defines the domain of external affairs in the Nash equilibrium. Hence, international agreements that are targeted directly to addressing the international externality can solve the policy inefficiencies associated with the externality while avoiding intrusion on the internal affairs of either country. In this sense, Proposition 3 highlights an important distinction between international agreements that mitigate international externalities and international agreements that erode national sovereignty, and points out that it can be possible to have the former without the latter.

We next characterize the violation of national sovereignty that occurs in the benchmark model when international agreements involve direct commitments over elements of $\mathbf{i}(g, x)$ and $\mathbf{i}^*(g^*, x)$. In particular, we will show that the indirect encroachment on sovereignty associated with such commitments can be more extensive than the direct violations themselves.

When international agreements involve direct commitments over elements of $\mathbf{i}(g, x)$ and/or $\mathbf{i}^*(g^*, x)$, a country's sovereignty may be violated in two ways according to our definition. First, sovereignty is violated whenever a government makes such commitments in an international

agreement, since the elements of $\mathbf{i}(g, x)$ and $\mathbf{i}^*(g^*, x)$ are matters that concern its internal affairs. In what follows, we say that the home (foreign) country's sovereignty over a policy instrument in $\mathbf{i}(g, x)$ ($\mathbf{i}^*(g^*, x)$) is *violated directly* by an international agreement whenever limits on this policy instrument are determined directly as a result of international negotiations. But a more subtle violation of sovereignty may also occur, if direct international commitments over elements of $\mathbf{i}(g, x)$ and/or $\mathbf{i}^*(g^*, x)$ result in unilateral choices over the remaining (non-negotiated) policy instruments that do not conform to the corresponding elements of $\hat{\mathbf{i}}(g, x)$ and/or $\hat{\mathbf{i}}^*(g^*, x)$, and therefore do not reflect the normal operation of a country's domestic institutions in the domain of its internal affairs. We will say that the home (foreign) country's sovereignty over a policy instrument is *violated indirectly* by an international agreement whenever the country's sovereignty over this instrument is not violated directly by the international agreement but the government's unilateral choice for this policy instrument differs from the corresponding element of $\hat{\mathbf{i}}(g, x)$ ($\hat{\mathbf{i}}^*(g^*, x)$) evaluated at the level of (g, x) ((g^*, x)) delivered under the agreement. Finally, we say that a country's sovereignty over a policy instrument is *violated (preserved)* whenever its sovereignty is violated directly or indirectly (neither directly nor indirectly).

As we next establish, direct violations of sovereignty typically imply further indirect violations as well, a feature that generally prevents governments from containing violations of sovereignty caused by international agreements to narrow subsets of policy instruments. In fact, as we now establish, this "contamination effect" implies that when an international agreement involves direct commitments over even a single element of $\mathbf{i}(g, x)$ or $\mathbf{i}^*(g^*, x)$, the number of policy instruments for which the agreement violates the associated country's sovereignty must be at least as great as the number of policy instruments for which the country's sovereignty is preserved under the agreement.

To establish this, we first introduce a notion of "interrelatedness" between policies. Taking the perspective of the domestic government, and recalling that L denotes the Step-1 Lagrangean for Program 1', we say that two policies u and v are *interrelated* if $L_{uv} \neq 0$ when L is evaluated at the maximized Step-1 choices $\hat{\mathbf{i}}(g, x)$. In words, when u and v are interrelated, a change in v alters the level of u preferred by the domestic government for delivering a given level of contribution g to the (given level of the) externality variable x . An exactly analogous interpretation applies for the foreign government.

We may now state:

Proposition 4. An international agreement that specifies levels for a subset of the elements of

$i(g, x)$ and $i^*(g^*, x)$ must (generically), for each country, violate that country's sovereignty over at least as many policy instruments as it preserves, provided that: (i) the agreement specifies at least one policy instrument for each government at a level different from its best-response level; and (ii) all policies are interrelated.

Proof: Consider an international agreement that specifies the levels for a subset of the elements of $\mathbf{i}(g, x)$ and $\mathbf{i}^*(g^*, x)$. We adopt the perspective of the domestic government. Let the elements of $\mathbf{i}(g, x)$ that are not determined directly by the international agreement be contained in the set \mathcal{H} . If \mathcal{H} is empty, then it is immediate that the statement of the proposition is satisfied, since in this case the sovereignty over all home-country instruments is violated (directly). If instead \mathcal{H} is non-empty, then the proposition is proved if it can be established that, to preserve the sovereignty of m home-country policy instruments, at least m home-country policy instruments must be directly negotiated (and therefore the home country's sovereignty over these instruments is violated directly). Let \mathbf{h} be the vector of non-negotiated home-country policies, and let $\bar{\mathbf{n}}$ be the vector of negotiated home-country policies whose levels are specified by the international agreement. Given any foreign policies \mathbf{i}^* , the home government's unilateral best-response choice of \mathbf{h} must solve the program:

$$\mathbf{Program\ 4:}\quad \text{For a given } \mathbf{i}^*, \max_{\mathbf{h}} G(\mathbf{h}, \bar{\mathbf{n}}, \tilde{x}(\mathbf{h}, \bar{\mathbf{n}}, \mathbf{i}^*)).$$

The first-order conditions for Program 4 are given by the analogue of (2.2) for the home government's instrument choices contained in \mathcal{H} . Now consider the partition of this program into the alternative two-step program:

$$\begin{aligned} \mathbf{Program\ 4'} : \quad & \text{Step 1. For a given } (g, x) : \max_{\mathbf{h}} G(\mathbf{h}, \bar{\mathbf{n}}, x) \\ & \text{s.t. } [g(\mathbf{h}, \bar{\mathbf{n}}, x) - g] = 0. \\ & \text{Step 2. For a given } \mathbf{i}^* : \max_{g, x} Q(\hat{\mathbf{h}}(g, x, \bar{\mathbf{n}}), \bar{\mathbf{n}}, g, x) \\ & \text{s.t. } f(g, g^*(\mathbf{i}^*, x), x) = 0, \end{aligned}$$

where $\hat{\mathbf{h}}(g, x, \bar{\mathbf{n}})$ is the solution from Step 1 and Q is the Step-1 Lagrangean. Arguments identical to those in the proof of Lemma 1 establish that Program 4' is a partition of Program 4. Hence, to complete the proof we need only observe that: (a) preserving the sovereignty of m home-country policy instruments requires that, with (g, x) set at the level delivered under the agreement, it must be possible to satisfy the Step-1 first-order conditions when evaluated at the

corresponding m elements of $\hat{i}(g, x)$, with $\hat{i}(g, x)$ itself evaluated at the level of (g, x) delivered under the agreement; and (b) with all policies interrelated, this in turn requires (generically) that there exist at least m policy instruments that are directly negotiated and can be used to “target” m of these Step-1 first-order conditions.¹¹ The only exception to this requirement occurs if the agreement fails to specify at least one policy instrument for each government at a level different from its best-response level, an exception that is ruled out by the condition (i) of the proposition. An analogous argument applies to the foreign government. **QED**

Hence, according to Proposition 4, if governments negotiate international agreements that encroach on the domain of their internal affairs with direct commitments over policy instruments in $\mathbf{i}(g, x)$ and $\mathbf{i}^*(g^*, x)$, national sovereignty will be violated, and the indirect erosion of sovereignty induced by these direct commitments may be greater and more far-reaching than the loss of sovereignty caused by the direct commitments themselves.

Moreover, Proposition 4 implies that, if national sovereignty has been violated in an international agreement by a direct commitment, the loss of sovereignty is not necessarily monotonic in the “depth” of the agreement, when depth is measured by the number of policy instruments in the domain of each country’s internal affairs that the agreement subjects to direct commitments. This is because, once the first commitment that falls within the domain of a country’s internal affairs has been added to an international agreement and hence violates directly the country’s sovereignty, the indirect violations of the country’s sovereignty induced by this direct commitment can be wide-spread (possibly covering *all* of the country’s policy instruments), and the introduction of further commitments in the agreement on policies that fall within the domain of the country’s internal affairs can then *reduce* the overall (direct plus indirect) violations of national sovereignty if those commitments reduce the number of policy instruments over which the country’s sovereignty is violated *indirectly* by the agreement. We record this in:

Corollary 1. If national sovereignty has been violated in an international agreement by a direct commitment, the loss of sovereignty is not necessarily increasing in the depth of the agreement.

¹¹The qualifier to generic cases in Proposition 4 refers to the possibility that optimal unilateral choices for some instruments might by chance happen to correspond to the policy levels needed for “targeting” other first-order conditions as described in the proof of the proposition.

5. Sovereignty and International Agreements in an Extended Model

In the previous section we established that, in principle, countries need not confront a choice between preserving their sovereignty and attaining international efficiency when they negotiate international agreements (Proposition 3). This is a striking result, derived within our benchmark model of section 2, and important questions remain as to (i) whether the result is likely to extend to environments beyond those captured by our benchmark model, and (ii) the degree to which prominent channels of interdependence between countries can be identified that find representation in the environment described by our benchmark model. We devote this and the next section to providing answers to these two questions.

As we noted in section 2, the structure in (2.1) rules out the possibility that $f_g g_{i_k} = 0$ for all k and/or that $f_{g^*} g_{i_m}^* = 0$ for all m , and thereby excludes cases where the externality variable x cannot be impacted by the policy instruments of some government. The question we consider in this section is how our results are impacted when the possibility ruled out by (2.1) arises. To this end, we now consider several extensions of the benchmark model that do not conform to the structure imposed by (2.1).

5.1. Small countries

An important possibility that is excluded from our benchmark model is that a country might be “small” in relation to its contribution to the externality variable x : if the home country is small in this sense, it would imply that $g_{i_k} = 0$ for all k (and hence $f_g g_{i_k} = 0$ for all k), while if the foreign country is small this would imply $g_{i_m}^* = 0$ for all m (and hence $f_{g^*} g_{i_m}^* = 0$ for all m).

We now consider an extension of the benchmark model in which there are many home and many foreign countries, each of which individually is small in the sense described just above. Formally, we subdivide the territory of the home country into N home countries and now assume that the overall contribution of the home countries to the externality is given by

$$g(\mathbf{i}^1, \dots, \mathbf{i}^N, x) \equiv \frac{1}{N} \sum_{j=1}^N g^j(\mathbf{i}^j, x).$$

Similarly, we subdivide the territory of the foreign country into N foreign countries and now

assume that the overall contribution of the foreign countries to the externality is given by

$$g^*(\mathbf{i}^{*1}, \dots, \mathbf{i}^{*N}, x) \equiv \frac{1}{N} \sum_{j=1}^N g^{*j}(\mathbf{i}^{*j}, x).$$

In this setting, with $\frac{\partial g^j}{\partial i_k^j}$ and $\frac{\partial g^{*j}}{\partial i_k^{*j}}$ assumed finite and bounded, we then have

$$\begin{aligned} \frac{\partial g}{\partial i_k^j} &= \frac{1}{N} \frac{\partial g^j}{\partial i_k^j} \longrightarrow 0 \text{ as } N \longrightarrow \infty \text{ for all } k \text{ and for } j \in \{1, \dots, N\}, \text{ and} \\ \frac{\partial g^*}{\partial i_m^{*j}} &= \frac{1}{N} \frac{\partial g^{*j}}{\partial i_m^{*j}} \longrightarrow 0 \text{ as } N \longrightarrow \infty \text{ for all } m \text{ and for } j \in \{1, \dots, N\}, \end{aligned}$$

and so for all $j \in \{1, \dots, N\}$ and for $N \longrightarrow \infty$ we have $f_g g_{i_k}^j = 0$ for all k and $f_{g^*} g_{i_m}^{*j} = 0$ for all m . An immediate implication is then that, as N goes to infinity, no single government acting alone can alter x with its policy choices: that is, $\tilde{x}_{i_k^j} = 0$ and $\tilde{x}_{i_k^{*j}} = 0$ for all j and all k . For simplicity, we assume that each home country is identical, and that each foreign country is identical, so that we may refer to “representative” home and foreign governments, but this is not essential for our results. All other features of the benchmark model remain the same.

It is direct to confirm two important implications of this extension of the benchmark model. First, suppressing the j superscript, the Nash policy choices of a representative home and foreign government solve $G_{i_k} = 0$ for $k = 1, \dots, I$ and $G_{i_k}^* = 0$ for $k = 1, \dots, I^*$. And second, letting $\boldsymbol{\iota}$ denote the vector of policies of all home governments $[\mathbf{i}^1 \mathbf{i}^2 \dots \mathbf{i}^N]$ and $\boldsymbol{\iota}^*$ denote the vector of policies of all foreign governments $[\mathbf{i}^{*1} \mathbf{i}^{*2} \dots \mathbf{i}^{*N}]$, and with \mathbf{i}^j the vector of policies of the j^{th} home government and with $\boldsymbol{\iota}^{-j}$ the vector of policies of all home governments other than j , $[\mathbf{i}^1 \dots \mathbf{i}^{j-1} \mathbf{i}^{j+1} \dots \mathbf{i}^N]$, and finally with $g^{-j}(\boldsymbol{\iota}^{-j}, x) \equiv g(\boldsymbol{\iota}, x) - g^j(\mathbf{i}^j, x)$, the minimal partition of the representative home government’s best-response policy choice problem is:

$$\begin{aligned} \mathbf{Program\ S} : \quad & \text{Step 1. For a given } x : \max_{\mathbf{i}^j} G(\mathbf{i}^j, x) \\ & \text{Step 2. For a given } \boldsymbol{\iota}^{-j} \text{ and } \boldsymbol{\iota}^* : \max_x Z(\hat{\nu}^j(x), x) \\ & \text{s.t. } f(g^j(\mathbf{i}^j, x) + g^{-j}(\boldsymbol{\iota}^{-j}, x), g^*(\boldsymbol{\iota}^*, x), x) = 0, \end{aligned}$$

where $\hat{\nu}^j(x)$ is the solution from Step 1 and Z is the Step-1 Lagrangean. The Step-2 determination of x is a trivial choice problem in Program S, because x is fully determined by the constraint once $\boldsymbol{\iota}^{-j}$ and $\boldsymbol{\iota}^*$ are given owing to the small size of country j . An analogous Program S* defines the minimal partition of the representative foreign government’s best-response policy choice problem.

According to Programs S and S*, each country's contribution to the externality variable x is a matter of its internal affairs when this contribution is vanishingly small. To see why, consider home government j . As compared to the minimal partition \mathcal{P}_0 described by Program 1' in the benchmark (large-country case) model, the key difference in the case of small countries is that g and hence f is unaffected by the choice of g^j , and so the home-government j 's choice of g^j can be moved from Step 2 to Step 1 without affecting the validity of the partition. This results in a minimal partition for the small-country case – described by Program S – that treats the choice of g^j as the internal affairs of home country j . An analogous interpretation applies for foreign country j . In essence the externality variable x is completely out of government j 's control owing to the vanishingly small impact of its contribution g^j , and government j has sole authority over everything else that matters for its objective function. As a consequence, when all countries are small, any international agreement that moves governments from their Nash policy choices must violate their sovereignty.

Together, these two implications lead to the following:

Proposition 5. When all countries are small in relation to their contribution to the externality variable x , attaining international efficiency is consistent with maintaining national sovereignty if and only if (i) governments are not mutually interdependent, or (ii) governments are mutually interdependent and $sign[G_x] \neq sign[G_x^*]$ when evaluated at the Nash policy choices. Under either condition (i) or condition (ii), the harmony between international efficiency and national sovereignty is preserved when small countries avoid making international commitments.

Proof: As implied by Programs S and S*, when all countries are small in relation their contribution to the externality variable x , sovereignty demands that each government be left to select its Nash policy choices. Hence, attaining international efficiency is consistent with maintaining sovereignty in the small-country case if and only if the associated Nash policy choices are efficient. If governments are not mutually interdependent, then arguments analogous to the proof of Proposition 1 establish that the Nash policy choices are efficient in the small-country case. Suppose, then, that governments are mutually interdependent. Evaluating the conditions for efficiency at the Nash policy choices of a representative home- and foreign-government, defined by $G_{i_k} = 0$ for $k = 1, \dots, I$ and $G_{i_k^*}^* = 0$ for $k = 1, \dots, I^*$, it is direct to confirm that (2.4)-(2.6) and the bottom line of (2.7) are satisfied, while the top line of (2.7) is satisfied if and only if $sign[G_x] \neq sign[G_x^*]$. QED

The interesting feature of Proposition 5 relates to the condition in part (ii). According to this condition, whether or not the apparent harmony between national sovereignty and international efficiency identified in Proposition 3 survives in a world of small countries hinges on whether governments agree or disagree in the Nash equilibrium over the direction that they would like the externality variable x to move. If all governments agree, then the Nash equilibrium in the small-country case is inefficient and an international agreement will be required to reach the efficiency frontier, implying necessarily that national sovereignty and international efficiency will stand in conflict in this case. However, if there is disagreement, then the Nash equilibrium in the small-country case is efficient, and in this case the harmony between national sovereignty and international efficiency identified in Proposition 3 survives in a world of small countries, though in this case as with condition (i) small countries must avoid international commitments to preserve this harmony. Of course, which of these two cases is applicable will depend on the nature of the externality variable x under consideration, but as we demonstrate in the next section, the latter case has special significance in the context of international trade agreements.¹²

5.2. Authority over an externality variable

At the opposite extreme of the small-country case considered above is the possibility that one country's policy choices completely determine the externality variable, so that either $f_g = 0$ or $f_g^* = 0$, a possibility that is ruled out in our benchmark model by (2.1). This possibility might arise, for example, if the foreign government determines the degree of religious freedom within its borders and the home-country government welfare is negatively impacted by religious persecution wherever it occurs.¹³

¹²To state Proposition 5 we have adopted a binary distinction between the limiting case of small countries that individually make a zero contribution to the externality variable and “non-small” countries that individually make a non-zero contribution to the externality variable. If countries were defined to be non-small only when their individual contributions to the externality variable exceeded some de-minimis level, then countries falling below this de-minimis level could not accept international commitments without violating their national sovereignty, and the potential conflicts between achieving international efficiency and maintaining national sovereignty described in Proposition 5 would be amplified.

¹³This possibility could also include, for example, the sorts of actions described by Jackson (2003, p 790) that might have been included in the original definition of Westphalian sovereignty (“... the nation-state’s power to violate virgins, chop off heads, arbitrarily confiscate property, torture citizens, and engage in all sorts of other excessive and inappropriate actions”). As we discussed above in note 4, Jackson’s approach is to update the definition of sovereignty so that there is no tradeoff between sovereignty and international intervention to address these kinds of actions. Our approach is to not tailor our definition of sovereignty on a case-by-case basis so that tradeoffs between sovereignty and international efficiency never exist, but rather to point out where such tradeoffs will be unavoidable, a result that is reported for this case in our Proposition 6 below.

To capture this possibility, let us suppose that the objectives of the home and foreign governments are now represented by the respective functions $G(\mathbf{i}, \tilde{x}(\mathbf{i}, \mathbf{i}^*), \tilde{y}(\mathbf{i}^*))$ and $G^*(\mathbf{i}^*, \tilde{x}(\mathbf{i}, \mathbf{i}^*), \tilde{y}(\mathbf{i}^*))$, with the equilibrium level of the externality variable $\tilde{x}(\mathbf{i}, \mathbf{i}^*)$ determined as before according to (2.1) whereas the equilibrium level of the externality variable $\tilde{y}(\mathbf{i}^*)$ is completely determined by \mathbf{i}^* , the policy vector of the foreign government. And let us suppose further that the home and foreign countries are mutually interdependent with respect to both x and y , so that $G_x \neq 0$ and $G_x^* \neq 0$ and also $G_y \neq 0$ and $G_y^* \neq 0$ at the Nash policy choices. It is easy to show that the Nash policy choices of the two governments are internationally inefficient, and in this case *each* of the externality variables is a source of inefficiency in the Nash equilibrium.

The minimal partition of the home government's best-response choice problem is now

Program A :

$$\begin{aligned} \text{Step 1. For a given } (g, x, y) : \quad & \max_{\mathbf{i}} G(\mathbf{i}, x, y) \\ & \text{s.t. } [g(\mathbf{i}, x) - g] = 0. \\ \text{Step 2. For a given } \mathbf{i}^* : \quad & \max_{g, x} L(\hat{v}(g, x, y), g, x, \tilde{y}(\mathbf{i}^*)) \\ & \text{s.t. } f(g, g^*(\mathbf{i}^*, x), x) = 0, \end{aligned}$$

where $\hat{v}(g, x, y)$ is the solution from Step 1 and L is the Step-1 Lagrangean. Program A implies that g , x and y comprise the home country's external affairs and hence could be subjected to constraints in an international agreement without violating the home country's sovereignty.

By contrast, the minimal partition of the foreign government's best-response choice problem is given by the following program:

Program A* :

$$\begin{aligned} \text{Step 1. For a given } (g^*, x) : \quad & \max_{\mathbf{i}^*} G^*(\mathbf{i}^*, x, \tilde{y}(\mathbf{i}^*)) \\ & \text{s.t. } [g^*(\mathbf{i}^*, x) - g^*] = 0. \\ \text{Step 2. For a given } \mathbf{i} : \quad & \max_{g^*, x} L^*(\hat{v}^*(g^*, x), g, x, \tilde{y}(\mathbf{i}^*)) \\ & \text{s.t. } f(g(\mathbf{i}, x), g^*, x) = 0, \end{aligned}$$

where $\hat{v}^*(g^*, x)$ is the solution from Step 1 and L^* is the Step-1 Lagrangean. Program A* implies that g^* and x – but not y – comprise the foreign country's external affairs. From this we may conclude that any commitment made by the foreign country in an international agreement beyond commitments to levels of g^* and/or x would violate its sovereignty. In particular, if national sovereignty is to be preserved, then the foreign government can make no commitments

over \mathbf{i}^* and hence $\tilde{y}(\mathbf{i}^*)$, indicating that this source of international inefficiency in the Nash equilibrium must then go unaddressed. This points to an unavoidable tension between national sovereignty and international efficiency in this case.

We summarize with:

Proposition 6. The harmony between national sovereignty and international efficiency breaks down whenever one country’s policies completely determine the level of an international externality variable.

5.3. A pure externality

A final case that we consider where the harmony between national sovereignty and international efficiency indicated by Proposition 3 can break down is when there exists an international externality variable that represents a “pure externality,” in the sense that the country generating the externality with its policy choices is not impacted by the externality variable. In our benchmark model with one externality variable, we allowed for this possibility (e.g., $G_x \neq 0 = G_x^*$) and showed that it implied an efficient Nash equilibrium under the conditions of our benchmark model (Proposition 1; see also note 2). But with multiple externalities, the required conditions for inefficiency of the Nash equilibrium are weakened, and the harmony between national sovereignty and international efficiency can be disrupted in the presence of pure externalities as a result. Here we consider a simple extension of our benchmark model to illustrate the point.

In particular, we introduce a second externality variable, y , whose determination can be characterized in an analogous fashion to the determination of x , but we assume that y is a concern only to the foreign government. And we now assume that the externality variable x is a concern only to the home government. That is, in this extension of the benchmark model, the objectives of the home and foreign governments are represented by the respective functions $G(\mathbf{i}, \tilde{x}(\mathbf{i}, \mathbf{i}^*))$ and $G^*(\mathbf{i}^*, \tilde{y}(\mathbf{i}, \mathbf{i}^*))$, with the two externality variables $\tilde{x}(\mathbf{i}, \mathbf{i}^*)$ and $\tilde{y}(\mathbf{i}, \mathbf{i}^*)$ defining the nature of the interdependence between the two countries. The new externality variable $\tilde{y}(\mathbf{i}, \mathbf{i}^*)$ is a pure externality imposed by the home government’s policy choices on the foreign country, and it could for example represent the level of water pollution which flows in one direction from the “upstream” home country (who is therefore not affected by the polluted water) to the “downstream” foreign country (who is affected by the polluted water). And similarly, we are assuming that the externality variables $\tilde{x}(\mathbf{i}, \mathbf{i}^*)$ is now also a pure externality imposed by the

foreign government's policy choices on the home country, and it could for example represent the level of air pollution which flows in one direction from the "upwind" foreign country (who is therefore not affected by the polluted air) to the "downwind" home country (who is affected by the polluted air). In analogy with the benchmark model and (2.1), the externality variables $\tilde{x}(\mathbf{i}, \mathbf{i}^*)$ and $\tilde{y}(\mathbf{i}, \mathbf{i}^*)$ are defined implicitly according to

$$\begin{aligned} f(g(\mathbf{i}, x), g^*(\mathbf{i}^*, x), x) &= 0, \text{ and} \\ c(q(\mathbf{i}, \mathbf{y}), q^*(\mathbf{i}^*, y), y) &= 0, \end{aligned}$$

where in analogy with the benchmark model we impose $f_g g_{i_k} \neq 0$ and now also $c_q q_{i_k} \neq 0$ for some k and $f_{g^*} g_{i_m}^* \neq 0$ and now also $c_{q^*} q_{i_m}^* \neq 0$ for some m . We also impose $G_x \neq 0$ and $G_y^* \neq 0$, but note that the home and foreign countries are not mutually interdependent with regard to either x or y (because $G_y = 0 = G_x^*$).

By construction, in this extension of the benchmark model the home government's best-response policy choice problem can be represented by Program 1' as in the benchmark model, and so it is direct to confirm that the home country's external affairs continue to be its choices over g and x , and do not include its choices over q and y . In an analogous fashion, it is direct to confirm that the foreign government's external affairs in this extended benchmark model are its choices over g^* and x , and also q^* and y .

In this setting, it is easy to see that a requirement for international efficiency is that the home government's policy instruments must be set so that the home government is indifferent to small changes in its policies that leave the externality variable y unaltered (since the foreign government would be indifferent to such changes). Formally, and letting $k = 1$ denote a domestic policy instrument for which $g_{i_k} \neq 0$ and $q_{i_k} \neq 0$, this condition may be written as:

$$\left[\frac{G_x f_g}{[f_g g_x + f_g^* g_x^* + f_x]} \right] \times [g_{i_1} g_{i_k}] \times \left[\frac{g_{i_k}}{g_{i_1}} - \frac{q_{i_k}}{q_{i_1}} \right] = q_{i_k} \times \left[G_{i_k} - \frac{g_{i_k} G_{i_1}}{g_{i_1}} \right] \text{ for } k = 2, 3, \dots, I. \quad (5.1)$$

But as we have just observed, the external affairs of the home government are limited to its choices over g and x . Consequently, the international commitments that the home government can take on without violating its sovereignty are limited to commitments over g and x , and result in unilateral choices over its instruments \mathbf{i} in light of any such commitments that solve Step 1 of Program 1' for a given (g, x) :

$$\begin{aligned} \max_{\mathbf{i}} G(\mathbf{i}, x) \\ \text{s.t. } [g(\mathbf{i}, x) - g] &= 0. \end{aligned}$$

The first order conditions associated with this problem are:

$$\left[G^{i_k} - \frac{g_{i_k} G^{i_1}}{g_{i_1}} \right] = 0 \quad \text{for } k = 2, 3, \dots, I. \quad (5.2)$$

Consider now the home government's Nash policy choices. These choices solve Step 1 of Program 1' for the Nash levels of (g, x) , and they must therefore satisfy (5.2). But these choices then violate (5.1), and must therefore be inefficient from an international perspective, unless

$$g_{i_k} \times \left[\frac{g_{i_k}}{g_{i_1}} - \frac{q_{i_k}}{q_{i_1}} \right] = 0 \quad \text{for } k = 2, 3, \dots, I, \quad (5.3)$$

which is to say unless all adjustments of the home-government policies that leave \tilde{x} unchanged also leave \tilde{y} unchanged. Intuitively, the home government will only be indifferent in the Nash equilibrium over small changes in its policies that leave \tilde{y} unchanged if these same policy changes leave \tilde{x} unchanged, because the home government cares about changes in \tilde{x} but does not care about changes in \tilde{y} . But in fact, since (5.2) must hold for the domestic choices of \mathbf{i} given any level of (g, x) , it follows that any international agreement that commits the home government to a level of (g, x) must fail to attain the international efficiency frontier unless (5.3) holds. With the sovereignty of the home government violated by any international agreement that goes further than commitments over g and x , and with (5.3) violated except by chance, a conflict between national sovereignty and international efficiency is effectively unavoidable.

We summarize with:

Proposition 7. The harmony between national sovereignty and international efficiency can break down in the presence of international externalities that take the form of pure externalities.

6. Sovereignty, International Trade Agreements and the GATT/WTO

We now return to our benchmark model and ask: Can prominent channels of interdependence between countries be identified that find representation in the environment described by our benchmark model? To some extent we have already answered this question in the affirmative, by offering in section 2 a number of specific illustrations of interdependencies that can be captured within the structure implied by (2.1). In this section we develop in detail the implications of our results in the particular context of international trade as a source of interdependence between countries. We begin in the next subsection by briefly reviewing the two-country two-good competitive general equilibrium trade model adapted to allow for the possibility of both

tariff and domestic regulatory policy choices as developed in Bagwell and Staiger (2001). We establish that this model is a special case of the benchmark model developed in section 2, and so all the results of sections 2-4 apply. We then show that when these results are interpreted within the context of our trade model, they indicate that the fundamental principles underlying GATT/WTO market access agreements offer a way to achieve internationally efficient policies without sacrificing national sovereignty, and that attaining international efficiency is consistent with maintaining national sovereignty in this setting regardless of whether (all) countries are big or small. We also extend our analysis of trade agreements to a multilateral setting, establish that agreement to abide by a nondiscrimination principle such as the GATT/WTO MFN rule does not violate a government's sovereignty, and identify a critical role for MFN if governments are to achieve internationally efficient policies without sacrificing national sovereignty when some (but not all) countries are small.¹⁴

6.1. Sovereignty in the Basic Two-Country Trade Model

We first describe the essential features of the two-country trade model of Bagwell and Staiger (2001). The home country exports good y to the foreign country in exchange for imports of good x . The local price of good x relative to good y in the home (foreign) country is denoted by p (p^*). The “world price” (i.e., relative exporter price or terms of trade) is denoted by p^w , and international arbitrage links each country's local price to the world price in light of its (non-prohibitive) tariff according to $p = \tau p^w \equiv p(\tau, p^w)$ and $p^* = p^w / \tau^* \equiv p^*(\tau^*, p^w)$, where τ (τ^*) is one plus the ad valorem import tariff of the home (foreign) country. In addition to its tariff, each country also imposes a vector of local regulations, \mathbf{r} (with length R) for the home country and \mathbf{r}^* (with length R^*) for the foreign country, that may impact local production and/or consumption decisions at given prices. Each country's vector of local regulations therefore acts as a vector of “shift” parameters in its import demand and export supply functions, and we assume that these functions are differentiable in their respective regulation levels.

Incorporating each country's vector of regulations into its import demand and export supply functions, we denote these functions for the home country by $M(\mathbf{r}, p, p^w)$ and $E(\mathbf{r}, p, p^w)$, respectively, and for the foreign country by $M^*(\mathbf{r}^*, p^*, p^w)$ and $E^*(\mathbf{r}^*, p^*, p^w)$, respectively. The

¹⁴The other important non-discrimination rule in the GATT/WTO is that of “national treatment,” which applies to non-border measures. In our formal model, the MFN rule would apply to tariffs, while the national treatment rule would apply to regulations. We focus here on the implications of the MFN rule for national sovereignty, but we conjecture that analogous findings could be formalized with regard to national treatment.

home and foreign budget constraints, which must hold for any p^w , may then be written as

$$p^w M(\mathbf{r}, p, p^w) = E(\mathbf{r}, p, p^w) \quad (6.1)$$

$$M^*(\mathbf{r}^*, p^*, p^w) = p^w E^*(\mathbf{r}^*, p^*, p^w). \quad (6.2)$$

The equilibrium world price, $\tilde{p}^w(\mathbf{r}, \tau, \mathbf{r}^*, \tau^*)$, is determined by the x -market-clearing requirement

$$M(\mathbf{r}, p(\tau, p^w), p^w) = E^*(\mathbf{r}^*, p^*(\tau^*, p^w), p^w), \quad (6.3)$$

where we have made explicit the dependence of the local prices on the tariffs and the world prices, and market clearing for good y is then implied by (6.1), (6.2) and (6.3).

Finally, we represent the objectives of the home and foreign governments with the general functions $W(\mathbf{r}, p, \tilde{p}^w)$ and $W^*(\mathbf{r}^*, p^*, \tilde{p}^w)$, respectively. These objective functions reflect an important assumption: governments care about the regulatory (and tariff) choices of their trading partners only because of the trade impacts of these choices (and therefore only because of the impacts of these choices on the equilibrium world price \tilde{p}^w). We assume that, holding its regulations and its local price fixed at levels that do not imply autarky, each government would prefer an improvement in its terms of trade,

$$\begin{aligned} W_{\tilde{p}^w}(\mathbf{r}, p, \tilde{p}^w) &< 0 \text{ for } M(\mathbf{r}, p, p^w) > 0, \text{ and} \\ W_{\tilde{p}^w}^*(\mathbf{r}^*, p^*, \tilde{p}^w) &> 0 \text{ for } M^*(\mathbf{r}^*, p^*, p^w) > 0. \end{aligned} \quad (6.4)$$

According to (6.4), governments like transfers of revenue from their trading partners. In the case of autarky, a change in the terms of trade holding its regulations and local price fixed should be irrelevant to a government, since there is no trade volume and continues to be no trade volume after the change, and so we assume as well that

$$\begin{aligned} W_{\tilde{p}^w}(\mathbf{r}, p, \tilde{p}^w) &= 0 \text{ for } M(\mathbf{r}, p, p^w) = 0, \text{ and} \\ W_{\tilde{p}^w}^*(\mathbf{r}^*, p^*, \tilde{p}^w) &= 0 \text{ for } M^*(\mathbf{r}^*, p^*, p^w) = 0. \end{aligned} \quad (6.5)$$

We leave government objectives otherwise unrestricted, and observe that these objectives are consistent with a wide variety of models of government behavior (see Bagwell and Staiger, 1999, 2002).

To establish that the trading environment we have just described is a special case of the benchmark model described in section 2, we define $\mathbf{i} \equiv [\mathbf{r} \ \tau]$ and $\mathbf{i}^* \equiv [\mathbf{r}^* \ \tau^*]$, and then define

$$\begin{aligned} G(\mathbf{i}, \tilde{p}^w) &\equiv W(\mathbf{r}, p(\tau, \tilde{p}^w), \tilde{p}^w) \\ G^*(\mathbf{i}^*, \tilde{p}^w) &\equiv W^*(\mathbf{r}^*, p^*(\tau^*, \tilde{p}^w), \tilde{p}^w) \end{aligned}$$

and define

$$\begin{aligned} m(\mathbf{i}, p^w) &\equiv M(\mathbf{r}, p(\tau, p^w), p^w) \\ m^*(\mathbf{i}^*, p^w) &\equiv M^*(\mathbf{r}^*, p^*(\tau^*, p^w), p^w). \end{aligned}$$

We then substitute (6.2) into (6.3) to rewrite the x -market-clearing requirement as

$$p^w m(\mathbf{i}, p^w) - m^*(\mathbf{i}^*, p^w) = 0. \quad (6.6)$$

It is now direct to confirm that (6.6) is a special case of (2.1) in which $x \equiv p^w$, $g(\mathbf{i}, x) \equiv m(\mathbf{i}, p^w)$, $g^*(\mathbf{i}^*, x) \equiv m^*(\mathbf{i}^*, p^w)$, and where

$$f(g(\mathbf{i}, x), g^*(\mathbf{i}^*, x), x) \equiv [xg(\mathbf{i}, x) - g^*(\mathbf{i}^*, x)],$$

with $p^w > 0$ and with the natural restrictions on the import demand functions $m(\mathbf{i}, p^w)$ and $m^*(\mathbf{i}^*, p^w)$ that $m_{i_k} \neq 0$ for some k and $m_{i_h}^* \neq 0$ for some h then ensuring that $f_g g_{i_k} \neq 0$ for some k and $f_{g^*} g_{i_h}^* \neq 0$ for some h , as required for the benchmark model under (2.1). In words, our two-country trade model is a special case of our benchmark model in which the international externality variable is the terms of trade, in which each country's contribution to the determination of the externality is the quantity of imports it demands at a given terms of trade, and in which these contributions are aggregated according to a market-clearing condition to determine the equilibrium level of the externality.

The minimal partition of the home government's best-response policy choice problem in the international trade setting, corresponding to Program 1' in the benchmark model, is then:

$$\begin{aligned} \mathbf{Program\ 1'} \text{ (Trade)} : \quad & \text{Step 1. For a given } (m, p^w) : \max_{\mathbf{i}} G(\mathbf{i}, p^w) \\ & \text{s.t. } [m(\mathbf{i}, p^w) - m] = 0. \\ & \text{Step 2. For a given } \mathbf{i}^* : \max_{m, p^w} L(\hat{i}(m, p^w), m, p^w) \\ & \text{s.t. } [p^w m - m^*(\mathbf{i}^*, p^w)] = 0, \end{aligned}$$

where $\hat{i}(m, p^w)$ is the solution from Step 1 and L is the Step-1 Lagrangean. The analogous minimal partition for the foreign government's best-response policy choice problem in the in-

ternational trade setting is:

$$\begin{aligned}
 \textbf{Program 1}^* \textbf{ (Trade)} : \quad & \textit{Step 1. For a given } (m^*, p^w) : \quad \max_{\mathbf{i}^*} G^*(\mathbf{i}^*, p^w) \\
 & \text{s.t. } [m^*(\mathbf{i}^*, p^w) - m^*] = 0. \\
 & \textit{Step 2. For a given } \mathbf{i} : \max_{m^*, p^w} L^*(\hat{i}^*(m^*, p^w), m^*, p^w) \\
 & \text{s.t. } [p^w m(\mathbf{i}, p^w) - m^*] = 0,
 \end{aligned}$$

where $\hat{i}^*(m^*, p^w)$ is the solution from Step 1 and L^* is the Step-1 Lagrangean.

Provided governments are mutually interdependent, which according to (6.4) and (6.5) will be the case in this setting if and only if they trade positive amounts in the Nash equilibrium, we may then conclude from Propositions 1-4 that, when the nature of interdependence across countries takes the form of international trade: (i) if governments negotiate commitments over policy instruments in \mathbf{i} and \mathbf{i}^* , their sovereignty will be violated, and the extent of the violation will in general not be limited only to those policy instruments that are directly negotiated; and (ii) governments may negotiate commitments over m , p^w and m^* which then hold the home (foreign) government to policy choices satisfying $[m(\mathbf{i}, p^w) - m] = 0$ ($[m^*(\mathbf{i}^*, p^w) - m^*] = 0$) and attain a position on the international efficiency frontier without violating their sovereignty.

Consider now what these findings suggest regarding the implication for national sovereignty of commitments negotiated in the GATT/WTO. In Bagwell and Staiger (2001), we observed that, when a government agrees to “bind” a tariff in a GATT/WTO negotiation, this government is not making a commitment that (i) holds it rigidly to its bound tariff level in the future and (ii) implies no obligations regarding future choices over its remaining policy instruments. Rather, the government is making a *market access commitment*, which is interpreted in the GATT/WTO as a commitment by the home (foreign) government to establish and maintain certain “conditions of competition for exporters into the domestic market.” Two important observations follow. First, a government can in principle fulfill a market access commitment with any combination of policy instruments that implies the agreed-upon conditions of competition for exporters into the domestic market.¹⁵ And second, a commitment to certain conditions of competition for exporters into the domestic market implies in turn a commitment to an

¹⁵We say “in principal” because, as pointed out in Bagwell and Staiger (2001), under current GATT/WTO practice the flexibility to fulfill market access commitments with any combination of policy instruments is not unlimited, and in particular would not be sufficient in certain circumstances to achieve internationally efficient outcomes in the way that we describe here. See also Bagwell, Mavroidis and Staiger (2002) for a more detailed discussion on this point and proposals that would introduce the required additional flexibility.

import level m (m^*) in the home (foreign) market when exporters price at p^w , and hence subjects the home (foreign) government to an implied constraint of the form $[m(\mathbf{i}, p^w) - m] = 0$ ($[m^*(\mathbf{i}^*, p^w) - m^*] = 0$) when making its policy choices. From these two observations, we may conclude that the fundamental commitments negotiated in the GATT/WTO are best interpreted as commitments over m , p^w and m^* rather than as commitments over policy instruments in \mathbf{i} and \mathbf{i}^* (see Bagwell and Staiger, 2002, for further elaboration on the interpretation of GATT/WTO market access commitments along these lines).¹⁶

Under this interpretation, then, Propositions 1-4 lend some formal support to the fundamental approach that rests at the heart of the GATT/WTO, in the sense that the negotiation of market access commitments can in principle allow governments to attain the international efficiency frontier without sacrificing national sovereignty.¹⁷ We summarize this finding in:

Proposition 8. A market access agreement between the home and the foreign government can achieve the international efficiency frontier without violating the sovereignty of either country.

Proposition 8 is related to Propositions 3 and 4 of Bagwell and Staiger (2001), but recasts these results from the perspective of the formal definitions of national sovereignty that we develop here.

Finally, we observe that the harmony between national sovereignty and international efficiency survives in a world of (all) small countries when the nature of interdependence across countries takes the form of international trade. To see this, note that by our definitions we have $G_{\tilde{p}^w} = \tau W_p + W_{\tilde{p}^w}$ and $G_{\tilde{p}^w}^* = (1/\tau^*)W_{p^*} + W_{\tilde{p}^w}^*$. When the two-country trade model is extended

¹⁶In interpreting GATT/WTO commitments as commitments over m , p^w and m^* , we are implicitly conditioning on economic fundamentals (preferences, technologies, endowments); but the same statement applies to our efficiency characterizations, so this seems the appropriate reference point.

¹⁷This is not to say that the GATT and now the WTO poses no threat to the national sovereignty of its member countries. Rather, our claim here is simply that market access commitments are consistent with the preservation of national sovereignty, and that the GATT/WTO can in principle steer clear of violations of the sovereignty of its members by adhering closely to the market access approach. Even then, in practice the process by which market access commitments are interpreted and enforced may place the sovereignty of members at risk in ways that are not captured by our formal modeling. For example, Keohane (2002, p. 8) points out that "...the classic conception of [Westphalian] sovereignty prohibits a government from agreeing to rules defining a process, over which it does not have a veto, that can confer obligations not specifically provided for in the original agreement," and Barfield (2001, pp. 42-69) argues that in practice the WTO Dispute Settlement Body may pose just such a threat to the Westphalian sovereignty of its members (though see also Keohane, 2002, p. 17, for a more qualified view). While such threats to sovereignty may indeed be real in practice, it is clear that in principle the WTO is designed not to pose such a threat: as Article 3.2 of the WTO Dispute Settlement Understanding states, "Recommendations and rulings of the [WTO Dispute Settlement Body] cannot add to or diminish the rights and obligations provided in the covered agreements."

to allow for many small home and foreign countries as in the analogous section-5.1 extension of our benchmark model, it is direct to show that the Nash equilibrium policy choices satisfy $W_p = 0$ ($W_{p^*}^* = 0$) for a representative home (foreign) government. Hence, when all countries are small we have $G_{\bar{p}^w} = W_{\bar{p}^w}$ and $G_{\bar{p}^w}^* = W_{\bar{p}^w}^*$ when evaluated at the Nash equilibrium. But using (6.4) and (6.5), Proposition 5 then implies that, when the nature of interdependence across countries takes the form of international trade, the harmony between national sovereignty and international efficiency survives in a world of (all) small countries: focusing on case (ii) of Proposition 5 where countries are mutually interdependent, the key point is that, in the international trade setting, governments generally disagree in the Nash equilibrium over the direction they would like the terms of trade (the externality variable) to move, as (6.4) indicates, and so in a world of small countries the Nash equilibrium is efficient and national sovereignty can be preserved at no cost with the avoidance of international commitments.

Together with Proposition 8, this last result is suggestive of a broad compatibility between maintaining national sovereignty and achieving international efficiency in trade matters. However, these findings fall short of establishing this claim in the multilateral setting in which real trade agreements (e.g., the GATT/WTO) operate, where some countries may be large, others small, and the non-discrimination (MFN) rule plays a prominent role. In the next section, we extend our two-country trade model to a three-country model in order to consider these issues.

6.2. MFN and Sovereignty in a Three-Country Trade Model

We now consider a three-country trade model analogous to Bagwell and Staiger (1999), extended to include standards. The home country exports good y to foreign countries 1 and 2 and imports good x from each of them. For simplicity, we do not allow trade between the two foreign countries, and so only the home country has the opportunity to set discriminatory tariffs across its trading partners. The local price of good x relative to good y in the home country (foreign country j) is denoted by p (p^{*j} , $j = 1, 2$). The “world price” (i.e., relative exporter price) for trade between the home country and foreign country j is denoted by p^{wj} , and international arbitrage links each country’s local price to the relevant world price in light of its (non-prohibitive) tariff according to $p = \tau^j p^{wj} \equiv p(\tau^j, p^{wj})$, and $p^{*j} = p^{wj} / \tau^{*j} \equiv p^{*j}(p^{wj}, \tau^{*j})$ for $j = 1, 2$, where τ^j (τ^{*j}) is one plus the ad valorem import tariff that the home country (foreign country j) applies to the imports from foreign country j (the home country). This

implies in turn that world prices are *linked* across bilateral relationships:

$$p^{w1} = \left[\frac{\tau^2}{\tau^1} \right] \times p^{w2}. \quad (6.7)$$

We note in particular that an MFN rule requires $\tau^1 = \tau^2 \equiv \tau$ and therefore implies $p^{w1} = p^{w2} \equiv p^w$ by (6.7). As in the two-country trade model above, in addition to its tariff, each country also imposes a vector of local regulations, \mathbf{r} (with length R) for the home country and \mathbf{r}^{*j} (with length R^{*j}) for foreign country j , that may impact local production and/or consumption decisions at given prices. Each country's vector of local regulations will therefore act as a vector of "shift" parameters in its import demand and export supply functions, and as before we assume that these functions are differentiable in their respective regulation levels. For future reference, we denote the home government's vector of policy instruments by $\mathbf{i} \equiv [\mathbf{r} \ \tau^1 \ \tau^2]$, and we denote the vector of policy instruments for foreign government j by $\mathbf{i}^{*j} \equiv [\mathbf{r}^{*j} \ \tau^{*j}]$.

Incorporating each country's vector of regulations into its import demand and export supply functions, we denote these functions for the home country by $M(\mathbf{r}, p, T)$ and $E(\mathbf{r}, p, T)$, respectively, and for foreign country j by $M^{*j}(\mathbf{r}^{*j}, p^{*j}, p^{wj})$ and $E^{*j}(\mathbf{r}^{*j}, p^{*j}, p^{wj})$, respectively, where T is the home-country's multilateral terms of trade, and is defined by

$$T(\mathbf{r}^{*1}, p^{*1}, \mathbf{r}^{*2}, p^{*2}, p^{w1}, p^{w2}) \equiv \sum_{k=1}^2 s^{*k}(\mathbf{r}^{*1}, p^{*1}, \mathbf{r}^{*2}, p^{*2}, p^{w1}, p^{w2}) \times p^{wk}$$

with

$$s^{*j}(\mathbf{r}^{*1}, p^{*1}, \mathbf{r}^{*2}, p^{*2}, p^{w1}, p^{w2}) \equiv \frac{E^{*j}(\mathbf{r}^{*j}, p^{*j}, p^{wj})}{\sum_{k=1}^2 E^{*k}(\mathbf{r}^{*k}, p^{*k}, p^{wk})} \quad \text{for } j = 1, 2.$$

Observe that an MFN rule requiring $\tau^1 = \tau^2$ implies $p^{w1} = p^{w2} = T \equiv p^w$ by (6.7). In any event, with T defined, the home and foreign budget constraints may then be written as

$$T \times M(\mathbf{r}, p, T) = E(\mathbf{r}, p, T), \quad \text{and} \quad (6.8)$$

$$M^{*j}(\mathbf{r}^{*j}, p^{*j}, p^{wj}) = p^{wj} E^{*j}(\mathbf{r}^{*j}, p^{*j}, p^{wj}) \quad \text{for } j = 1, 2. \quad (6.9)$$

The pair of equilibrium world prices, $\tilde{p}^{wj}(\mathbf{i}, \mathbf{i}^{*1}, \mathbf{i}^{*2})$ for $j = 1, 2$, are then determined by the linkage condition (6.7) together with the requirement of market clearing for good x ,

$$M(\mathbf{r}, p, T) = \sum_{k=1}^2 E^{*k}(\mathbf{r}^{*k}, p^{*k}, p^{wk}), \quad (6.10)$$

with market clearing for good y then implied by (6.8) and (6.9).

Finally, in analogy with the two-country trade model, we represent the objectives of the home and foreign government $j = 1, 2$ with the general functions $W(\mathbf{r}, p, T)$ and $W^{*j}(\mathbf{r}^{*j}, p^{*j}, \tilde{p}^{wj})$, respectively. As before, we assume that, holding its regulations and its local price fixed, and provided that its regulations and local price do not imply autarky, each government would prefer an improvement in its terms of trade:

$$W_T(\mathbf{r}, p, T) < 0 \text{ for } M(\mathbf{r}, p, T) > 0, \text{ and } W_{\tilde{p}^{wj}}^{*j}(\mathbf{r}^{*j}, p^{*j}, \tilde{p}^{wj}) > 0 \text{ for } M^{*j}(\mathbf{r}^{*j}, p^{*j}, p^{wj}) > 0.$$

For simplicity in this section we consider only the case where Nash policy choices do not imply autarky. We leave government objectives otherwise unrestricted.

In the Nash equilibrium of the three-country trade model, the home government chooses its best-response policies by solving

$$\mathbf{Program\ 5:} \text{ For a given } \mathbf{i}^{*1} \text{ and } \mathbf{i}^{*2}, \max_{\mathbf{i}} W(\mathbf{r}, p(\tau^j, \tilde{p}^{wj}), T),$$

at the same time that foreign government j , for $j = 1, 2$, chooses its best- response policies by solving

$$\mathbf{Program\ 5}^{*j}: \text{ For a given } \mathbf{i} \text{ and } \mathbf{i}^{*-j}, \max_{\mathbf{i}^{*j}} W^{*j}(\mathbf{r}^{*j}, p^{*j}(\tau^{*j}, \tilde{p}^{wj}), \tilde{p}^{wj}).$$

We next show that the Nash policy choices defined by the simultaneous solutions to Program 5 and Program 5^{*j} may be written in an equivalent form in which each government's best-response program is partitioned into a two-step choice problem.

Because foreign countries 1 and 2 each trade with only one partner (the home country), they each face a single international externality variable (\tilde{p}^{wj}), and so the minimal partition of Program 5^{*j} into a two- step choice problem is completely analogous to Program 1['], the minimal partition of the foreign country in the two-country trade model of the previous section. However, the international externality faced by the home country is more complicated, owing to the possibility that it trades with two trading partners at two different bilateral world prices.

Nevertheless, as we now demonstrate, an analogous 2-step partition can be developed for the home government's best-response problem. In particular, consider the following 2-step program

for the home government:

$$\begin{aligned}
\mathbf{Program\ 5'} : \quad & \text{Step 1. For a given } (M, T) : \max_{\mathbf{r}, p} W(\mathbf{r}, p, T) \\
& \text{s.t. } [M(\mathbf{r}, p, T) - M] = 0. \\
& \text{Step 2. For a given } \mathbf{i}^{*1} \text{ and } \mathbf{i}^{*2} : \max_{M, p^{w1}, p^{w2}} Y(\hat{\mathbf{r}}(M, T(\cdot)), \hat{p}(M, T(\cdot)), M, T(\cdot)) \\
& \text{s.t. } [M - \sum_{k=1}^2 E^{*k}(\mathbf{r}^{*k}, p^{*k}(\tau^{*k}, p^{wk}), p^{wk})] = 0,
\end{aligned}$$

where $\hat{\mathbf{r}}(M, T(\cdot))$ and $\hat{p}(M, T(\cdot))$ are the solutions from Step 1 and Y is the Step-1 Lagrangean, and where $T(\cdot)$ denotes $T(\mathbf{r}^{*1}, p^{*1}(p^{w1}, \tau^{*1}), \mathbf{r}^{*2}, p^{*2}(p^{w1}, \tau^{*1}), p^{w1}, p^{w2})$. Observe that a value for T is determined for given \mathbf{i}^{*1} and \mathbf{i}^{*2} once p^{w1} and p^{w2} are determined, and so the Step-2 choice problem over M , p^{w1} and p^{w2} determines a value for M and T , which are each taken as given in the Step-1 choice problem. We first state:

Lemma 4. Program 5' is a minimal partition of the home-government's best-response choice problem defined by Program 5.

Proof: See Appendix.

According to Lemma 4, we may conclude that the external affairs of the home country are its government's choices over M , p^{w1} and p^{w2} (and by implication M and T). In analogy with the two-country trade model, these choices can be interpreted as determining the level of market access that the home country affords to each of its trading partners (as defined by the volume of imports it would accept at a particular multilateral terms of trade). The matters that concern the home country's internal affairs are then its choice of G and its government's choices over $\mathbf{r}(M, T(\cdot))$ and $p(M, T(\cdot))$. However, observing that $p = \tau^1 p^{w1} = \tau^2 p^{w2}$, we may restate the home government's choice over $p(M, T(\cdot))$ equivalently as a choice over $\tau^1(M, T, p^{w1})$ and $\tau^2(M, T, p^{w2})$ where $\tau^1(M, T, p^{w1}) \equiv p(M, T)/p^{w1}$ and $\tau^2(M, T, p^{w2}) \equiv p(M, T)/p^{w2}$. Recalling now that the MFN rule requires $\tau^1 = \tau^2 \equiv \tau$ and hence $p^{w1} = p^{w2} = T \equiv p^w$ by (6.7), but that the MFN rule leaves the level of τ and therefore p unrestricted, it follows that the MFN rule places restrictions on the home country's external affairs (its Step-2 choices) but places no restrictions nor introduces any distortions in the home country's internal affairs (its Step-1 choices). We may therefore state:

Proposition 9. Abiding by the non-discrimination rule does not violate national sovereignty.

Proposition 9 reflects the following intuition. Discriminatory tariffs make possible certain market access choices that would be impossible under MFN. But market access (Step-2) choices are the external affairs of a country, and so restrictions can be placed on these choices through voluntary international agreement without violating national sovereignty. And given any market access choices that would be feasible under MFN, discriminatory tariffs do not create any additional possibilities for delivering these market access levels. This feature is reflected in the fact that the Step-1 choices of Program 5' may be expressed as choices over domestic regulations \mathbf{r} and the domestic price level p , and for these choices the MFN restriction has no bearing. Hence, a country's sovereignty is violated neither directly nor indirectly when it agrees to abide by the MFN rule.¹⁸

According to Proposition 9, an agreement to abide by the MFN rule in the three-country trade model does not violate the sovereignty of any country. But once the MFN rule is imposed, it is direct to show that the three-country trade model is a straightforward extension of the two-country trade model, and exhibits all the same properties. In particular, it then follows by Propositions 8 and 9 that a market access agreement between the home and foreign governments in the three-country trade model that also imposes an MFN requirement can achieve the international efficiency frontier without violating the sovereignty of any country. This falls short of the stronger claim that the MFN rule is *required* to make the attainment of internationally efficient outcomes compatible with the maintenance of national sovereignty. But as we next show, this stronger claim can be made when the three-country trade model is extended to allow that some (but not all) countries are small.

To see this, let us suppose that foreign country 2 in the three-country trade model is now decomposed into a region of small foreign countries as with our small-country extension in section 5.1, so that no single government k in foreign region 2 can, acting alone, alter \tilde{p}^{w2} with its policy choices. As we established in section 5.1, small countries must avoid international commitments in order to maintain their sovereignty, and so any international agreement that moves the governments in foreign region 2 from their best-response policy choices must violate their sovereignty. It is easy to see from Program 5*^j that the best-response policy choices of a representative government in foreign region 2 (suppressing the individual country superscript

¹⁸We also note that Proposition 9 is in line with Rabkin (2004, pp. 131-134) who, arguing from an historical and legal perspective, concludes that MFN obligations are consistent with the preservation of national sovereignty.

k) solve

$$W_p^{*2} = 0 \text{ and } W_{r_i}^{*2} = 0 \text{ for } i = 1, 2, \dots, R^{*2}. \quad (6.11)$$

The question, then, is whether the home government and foreign government 1 can undertake commitments that (i) do not violate their sovereignty and (ii) attain a position on the international efficiency frontier when the governments in foreign region 2 choose policies that satisfy (6.11).

We first establish:

Proposition 10. An international agreement can attain a point on the international efficiency frontier and satisfy (6.11) if and only if it satisfies the MFN rule.

Proof: See Appendix.

As we establish in the Appendix, to achieve international efficiency and satisfy (6.11), the home government and foreign government 1 must adopt policies that abide by the MFN rule and satisfy

$$W_p = 0 = W_p^{*1}, \quad W_{r_i} = 0 \text{ and } W_{r_i}^{*1} = 0 \text{ for } i = 1, 2, \dots, R^{*1}. \quad (6.12)$$

Referring to market access agreements that achieve the market access levels implied by (6.11), (6.12) and the MFN restriction as *non-discriminatory politically optimal market access agreements*, and utilizing Propositions 8-10, we may now state:

Proposition 11. If some (but not all) countries are “small,” then achieving international efficiency and preserving national sovereignty are mutually consistent goals of an international agreement if and only if the agreement satisfies the MFN requirement. In particular, non-discriminatory politically optimal market access agreements provide the unique path to achieving international efficiency while preserving national sovereignty in this setting.

In effect, if small countries are asked to make market access commitments, their sovereignty will be violated. If this is to be avoided, then small countries must be left unconstrained to choose their best-response policies in any international agreement. This requirement, though, is consistent with international efficiency only when tariffs also conform to the MFN requirement, as indicated by Proposition 10. As a consequence, in an international trade setting Proposition 11 suggests that a non-discrimination rule is “complementary” to preserving the national sovereignty of small countries in the following sense: the sovereignty of small countries can

be preserved under an internationally efficient agreement only if that agreement abides by the MFN requirement. More broadly, and in light of our finding in Proposition 9 that the MFN requirement itself involves no compromise of national sovereignty, our three-country results suggest that a non-discrimination rule coupled with a market access agreement can facilitate the attainment of internationally efficient outcomes that do not compromise national sovereignty.

6.3. Sovereignty, GATT and the WTO

When viewed together, the results from the previous two sub-sections have potentially important implications for the design of the WTO and its predecessor, the GATT. The GATT/WTO has from its inception been concerned most fundamentally with non-discriminatory market access commitments, and it has traditionally sought to anchor these commitments with negotiations over border measures (e.g., tariffs) that are “multilateralized” through the MFN requirement. But this tradition is being eroded on two fronts. First, the extent and importance of discriminatory trade agreements (permitted by GATT/WTO exceptions to its MFN requirement) has increased dramatically in recent decades. And second, increasingly the WTO is thought of as a potential forum for the negotiation of international commitments on a host of non-border policies that are deemed to have important market access consequences, ranging from labor standards to environmental regulations to domestic subsidies to competition policy. Our results highlight the fundamental implications of these developments for the potential conflicts between international efficiency and national sovereignty within the WTO. Specifically, as these results indicate, the further the WTO departs from facilitating agreements that take the form of non-discriminatory market access commitments, the more it is likely to pose a (direct and indirect – and in principle, unnecessary) threat to the sovereignty of its member countries.

7. Conclusion

What are the sovereign rights of nations in an interdependent world, and to what extent do these rights stand in the way of achieving internationally efficient outcomes? In this chapter, we have proposed answers to these two questions. Our answers, of course, depend on the definition of national sovereignty. Sovereignty is a complex concept with many features, and any definition has advantages and disadvantages. Our approach is to formalize the Westphalian norm of “non-intervention in the internal affairs of other states” in a way that we believe captures several key features of Westphalian sovereignty emphasized in the international political econ-

omy literature, features that seem especially relevant in the context of voluntary international agreements. An important advantage of this approach is that it is analytically tractable. Using this formalization, we show how Nash choice problems can be partitioned in a way that allows a characterization of the degree and nature of sovereignty that governments possess in the Nash equilibrium. This characterization, in turn, provides a benchmark from which to formally assess the implications for national sovereignty of international agreements of various designs.

In the context of our benchmark model of international interdependence, we find that in principle there is no inherent conflict between the attainment of international efficiency through international agreements and preserving national sovereignty. As our benchmark model is general enough to cover channels of international interdependence that can take a variety of forms, we view this finding as pointing to important possibilities for eliminating conflicts between international efficiency and national sovereignty through appropriate design of international agreements. Our findings here highlight an important distinction between international agreements that mitigate international externalities and international agreements that erode national sovereignty, and point out that it can be possible to have the former without the latter.

In the particular case of international trade relations, which we argue is captured well by our benchmark model, we find that a number of the foundational aspects of the GATT/WTO, such as its emphasis on market access commitments and the MFN rule, offer a possible means of achieving international efficiency without eroding national sovereignty. In this regard, we give formal support to the observation of Rabkin (1998):

“Probably the single most effective and consequential international program of the postwar era has been the mutual reduction of trade barriers under the General Agreement on Tariffs and Trade, initiated in 1947. Reasonable questions may be raised about certain aspects of the World Trade Organization, established in 1995 to help administer GATT norms. But, fundamentally, the trading system is quite compatible with traditional notions of sovereignty. It was developed on the foundations of much older sorts of international agreement, which would have been quite recognizable to the Framers of the Constitution.” Rabkin, pp. 85-86.

However, our results suggest that the maintenance of this compatibility depends crucially on being true to these fundamental principles: the further away the WTO moves from a market-access focus and adherence to MFN, the more likely will conflicts arise within the WTO between

international efficiency and national sovereignty.

Importantly, we also find that the harmony in our benchmark model between international efficiency and national sovereignty is not always present in the universe of international relations among national governments. Rather, as the various extensions to our benchmark model illustrate, this harmony depends on the structure of the international externalities that define the nature of international interdependence and that give rise to the policy inefficiencies under non-cooperative policy setting. Together with our other findings, this last point reinforces the importance of understanding the nature of the international externalities that are the source of the problem for an international agreement to solve, so that those tensions between sovereignty and efficiency that are unnecessary can be avoided, while those tensions that are inescapable can be anticipated and minimized through careful institutional design.

Where tensions between national sovereignty and international efficiency are inescapable according to our formal analysis, it may be that further refinements to the notion of national sovereignty, perhaps tailored to the case at hand, could help to ease these tensions. A key message of our paper, however, is that such refinements, to the extent they are possible, can be guided by our formal approach and instructive findings. In this way, our analysis provides a framework for further systematic exploration into notions of sovereignty and their implications for efficiency.

At the same time, in circumstances where a tension between national sovereignty and international efficiency is inescapable, our formal analysis leaves unanswered an important question: in such circumstances, is the preservation of national sovereignty worth the sacrifice of international efficiency? At a general level, there are two ways that this question might be approached.

First, if as we assume the general government objectives in our formal analysis capture all relevant considerations for the associated countries, then there is a clear case for governments to pursue international efficient agreements, even if national sovereignty as we define that term is violated. From this perspective, the contribution of our analysis is to identify settings in which international efficiency can be achieved without violating national sovereignty.

Second, if there are explicit costs associated with the erosion of national sovereignty that are not captured in our formal analysis, then a formal answer to this question would require introducing those costs into the model. For example, such costs might arise if the quality of information depreciates with the distance between the decision-maker and those most directly impacted by the decision. A complete analysis would then weigh such costs against the efficiency

benefits that an international agreement that violates national sovereignty could make possible. Introducing such micro-foundations for concerns about violations of sovereignty is an important area for further research.

8. References

Bagwell, Kyle, Petros C. Mavroidis and Robert W. Staiger, "It's a Question of Market Access," *The American Journal of International Law*, January 2002, pp. 56-76.

Bagwell, Kyle, and Robert W. Staiger, "An Economic Theory of GATT," *American Economic Review*, March 1999, pp. 215-248.

Bagwell, Kyle, and Robert W. Staiger, "Domestic Policies, National Sovereignty and International Economic Institutions," *Quarterly Journal of Economics*, May 2001, pp. 519-562.

Bagwell, Kyle, and Robert W. Staiger, *The Economics of the World Trading System*, MIT Press, 2002.

Bagwell, Kyle, and Robert W. Staiger, "National Sovereignty in an Interdependent World," NBER Working Paper No 10249, January 2004.

Barfield, Claude E., *Free Trade, Sovereignty, Democracy: The Future of the World Trade Organization*, The AEI Press, Washington D.C.. 2001.

de Vattel, Emmerich, *The Law of Nations*, Translated by Joseph Chitty, T.& J. W. Johnson & Co., Law Booksellers, Philadelphia. 1872.

Gray, John, "Introduction," in *John Stuart Mill: On Liberty and Other Essays*, Oxford World's Classics: Oxford University Press, Oxford. 1991.

Jackson, John H., "Sovereignty-Modern: A New Approach to an Outdated Concept," *American Journal of International Law* vol 97, 2003, pp. 782-802.

Keohane, Robert O., "Ironies of Sovereignty: The European Union and the United States," *JCMS: Journal of Common Market Studies*, vol 40, 2002, pp. 743-765.

Krasner, Stephen D., *Sovereignty: Organized Hypocrisy*, Princeton University Press. 1999.

Rabkin, Jeremy, *Why Sovereignty Matters*, The AEI Press, Washington, D.C.. 1998.

Rabkin, Jeremy, *The Case for Sovereignty*, The AEI Press, Washington, D.C. 2004.

9. Appendix

In this Appendix, we provide proofs of all lemmas and propositions that are not proved in the body of the paper.

Proposition 1: The Nash equilibrium of the benchmark model is inefficient if and only if $G_x \neq 0$ and $G_x^* \neq 0$ at the Nash policy choices.

Proof: Using (2.1), we may derive that

$$\tilde{x}_{i_k} = \frac{-f_g g_{i_k}}{f_g g_x + f_{g^*} g_x^* + f_x} \quad \text{for } k = 1, 2, \dots, I, \quad (\text{A1})$$

$$\tilde{x}_{i_k^*} = \frac{-f_{g^*} g_{i_k^*}}{f_g g_x + f_{g^*} g_x^* + f_x} \quad \text{for } k = 1, 2, \dots, I^*. \quad (\text{A2})$$

With (A1) and (A2), it is direct to show that the Nash conditions (2.2) and (2.3) imply efficiency conditions (2.4) and (2.5). Further, substituting the Nash conditions (2.2) and (2.3) into efficiency condition (2.6) yields

$$G_x^* \tilde{x}_{i_1^*} G_x \tilde{x}_{i_1} = 0,$$

which is violated at the Nash policy choices if and only if $G_x \neq 0$ and $G_x^* \neq 0$ at these policy choices. Finally, it may be confirmed that (2.7) will be satisfied at the Nash policy choices if either $G_x = 0$ or $G_x^* = 0$ at these policy choices. QED

Lemma 1: Program 1' is a partition of Program 1.

Proof: We prove this by establishing that the first-order conditions associated with Program 1' are equivalent to the first-order conditions associated with Program 1, given by (2.2). Letting λ_1 denote the Lagrange multiplier associated with the constraint in Step 1 of Program 1', and letting λ_2 denote the Lagrange multiplier associated with the constraint in Step 2 of Program 1', the first-order conditions associated with Step 1 are given by

$$G_{i_k} - \lambda_1 g_{i_k} = 0 \quad \text{for } k = 1, 2, \dots, I, \quad (\text{A3})$$

while the first order conditions associated with Step 2 are given by

$$\begin{aligned} \lambda_1 - \lambda_2 f_g &= 0, \text{ and} \\ G_x - \lambda_1 g_x - \lambda_2 [f_{g^*} g_x^* + f_x] &= 0. \end{aligned} \quad (\text{A4})$$

Eliminating λ_1 and λ_2 from (A3)-(A4) yields (2.2). QED

Lemma 4: Program 5' is a minimal partition of the home-government's best-response choice problem defined by Program 5.

Proof: We first establish that Program 5' is a partition of Program 5 (Part I). We then argue that it is a minimal partition (Part II).

Part I: The first-order conditions associated with Program 5 are

$$W_{r_i} + \tau^j W_p \times \frac{\partial \tilde{p}^{wj}}{\partial r_i} + W_T \times \frac{dT}{dr_i} = 0 \text{ for } i = 1, 2, \dots, R, \text{ and} \quad (\text{A5})$$

$$W_p + \theta^j W_T = 0 \text{ for } j = 1, 2, \quad (\text{A6})$$

where $\theta^j \equiv \frac{dT/d\tau^j}{dp/d\tau^j}$. Observe that by (6.7), $\tau^1[\partial \tilde{p}^{w1}/\partial r_i] = \tau^2[\partial \tilde{p}^{w2}/\partial r_i]$, and so (A5) may be equivalently evaluated for either $j = 1, 2$. The first-order conditions associated with Step 1 and Step 2 of Program 5', with γ_1 and γ_2 denoting the Lagrange multiplier on the constraints in Step 1 and Step 2, respectively, are

$$W_{r_i} + \gamma_1 M_{r_i} = 0 \text{ for } i = 1, 2, \dots, R, \quad (\text{A7})$$

$$W_p + \gamma_1 M_p = 0, \quad (\text{A8})$$

$$-\gamma_1 + \gamma_2 = 0, \quad (\text{A9})$$

$$\tau^1 W_p + \gamma_1 \tau^1 M_p + [W_T + \gamma_1 M_T] \times \left[\frac{1}{\tau^{*1}} \frac{\partial T}{\partial p^{*1}} + \frac{\partial T}{\partial p^{w1}} \right] - \gamma_2 \left[\frac{1}{\tau^{*1}} E_{p^{*1}}^{*1} + E_{p^{w1}}^{*1} \right] = 0, \text{ and} \quad (\text{A10})$$

$$[W_T + \gamma_1 M_T] \times \left[\frac{1}{\tau^{*2}} \frac{\partial T}{\partial p^{*2}} + \frac{\partial T}{\partial p^{w2}} \right] - \gamma_2 \left[\frac{1}{\tau^{*2}} E_{p^{*2}}^{*2} + E_{p^{w2}}^{*2} \right] = 0. \quad (\text{A11})$$

By using (A9) and (A11) to derive an expression for γ_1 and noting that $\frac{dT}{dp^{w1}} = \left[\frac{1}{\tau^{*2}} \frac{\partial T}{\partial p^{*2}} + \frac{\partial T}{\partial p^{w2}} \right]$, (A8) may be written as

$$W_p + \left[\frac{M_p \frac{dT}{dp^{w2}}}{E_{p^{w2}}^{*2} - M_T \frac{dT}{dp^{w2}}} \right] \times W_T = 0. \quad (\text{A12})$$

Next, we observe that (A6) implies $\theta^1 = \theta^2$, which can be manipulated to yield

$$\frac{\partial \tilde{p}^{w1}}{\partial \tau^2} = \left[\frac{\tilde{p}^{w2}}{\tilde{p}^{w1}} \right] \times \left[\frac{dT/dp^{w2}}{dT/dp^{w1}} \right] \times \left[\frac{\partial \tilde{p}^{w2}}{\partial \tau^1} \right],$$

which in turn allows θ^2 to be written as

$$\theta^2 = \left[\frac{dT}{dp^{w2}} \right] \times \frac{\left[\frac{\partial \tilde{p}^{w2}}{\partial \tau^2} \right] + \left[\frac{\tilde{p}^{w2}}{\tilde{p}^{w1}} \right] \left[\frac{\partial \tilde{p}^{w2}}{\partial \tau^1} \right]}{\tilde{p}^{w2} + \tau^2 \left[\frac{\partial \tilde{p}^{w2}}{\partial \tau^2} \right]}. \quad (\text{A13})$$

Using the linkage condition (6.7) and the market-clearing condition (6.10), expressions for $\frac{\partial \tilde{p}^{w2}}{\partial \tau^1}$ and $\frac{\partial \tilde{p}^{w2}}{\partial \tau^2}$ may be derived which, when substituted into (A13), yield

$$\theta^2 = \times \frac{M_p \times \left[\frac{dT}{dp^{w2}} \right]}{E_{p^{w2}}^{*2} - M_T \frac{dT}{dp^{w2}}}. \quad (\text{A14})$$

Therefore, by substituting (A14) into (A6) and observing that the resulting expression is identical to (A12), we may conclude that (A9), (A11) and (A8) imply (A6). Similarly, we use (A9) and (A10) to derive an alternative expression for γ_1 , which allows (A7) to be written as

$$W_{r_i} + \tau^1 W_p \times \left[\frac{M_{r_i}}{E_{p^{w1}}^{*1} - M_{p^{w1}}} \right] + W_T \times \left[\frac{dT}{dp^{w1}} \right] \times \left[\frac{M_{r_i}}{E_{p^{w1}}^{*1} - M_{p^{w1}}} \right] = 0. \quad (\text{A15})$$

Now using (6.7) and (6.10), we may derive that

$$\frac{\partial \tilde{p}^{w1}}{\partial r_i} = \left[\frac{M_{r_i}}{E_{p^{w1}}^{*1} - M_{p^{w1}}} \right]. \quad (\text{A16})$$

Substituting (A16) into (A15) yields an expression identical to (A5). Hence, we may conclude that (A9), (A10) and (A7) imply (A5).

Part II: The proof that Program 5' is a minimal partition of Program 5 proceeds in the same way as the proof of Lemma 3. QED

Proposition 10: An international agreement can attain a point on the international efficiency frontier and satisfy (6.11) if and only if it satisfies the MFN rule.

Proof: To prove this proposition, we first characterize the efficiency frontier of the 3-country model (that is, for notational simplicity, we treat the foreign governments in region *2 as if they were all identical, but this is not essential for the result). To this end, fix foreign welfare levels \bar{W}^{*j} for $j = \{1, 2\}$ and define $\tilde{p}^{wj}(\mathbf{r}^{*j}, \tau^{*j}, \bar{W}^{*j})$ implicitly by $W^{*j}(\mathbf{r}^{*j}, p^{*j}(\tau^{*j}, \tilde{p}^{wj}), \tilde{p}^{wj}) = \bar{W}^{*j}$ for $j = \{1, 2\}$. Observe that

$$\frac{\partial \tilde{p}^{wj}}{\partial \tau^{*j}} = \frac{p^{*j} W_{p^{*j}}^{*j}}{W_{p^{*j}}^{*j} + \tau^{*j} W_{\tilde{p}^{wj}}^{*j}}; \quad \text{and} \quad \frac{\partial \tilde{p}^{wj}}{\partial r_i^{*j}} = \frac{-\tau^{*j} W_{r_i^{*j}}^{*j}}{W_{p^{*j}}^{*j} + \tau^{*j} W_{\tilde{p}^{wj}}^{*j}}, \quad (\text{A17})$$

for $i = 1, 2, \dots, R^{*j}$ and $j = \{1, 2\}$. We may now define

$$\bar{T}(\{\mathbf{r}^{*j}\}, \{\tau^{*j}\}, \{\bar{W}^{*j}\}) \equiv T(\{\mathbf{r}^{*j}\}, \{p^{*j}(\tau^{*j}, \tilde{p}^{wj}(\mathbf{r}^{*j}, \tau^{*j}, \bar{W}^{*j}))\}, \{\tilde{p}^{wj}(\mathbf{r}^{*j}, \tau^{*j}, \bar{W}^{*j})\}),$$

and observe that, by the market-clearing condition (6.10), a value of p is implied, which we denote by $\bar{p}(\mathbf{r}, \mathbf{r}^{*j}, \tau^{*j}, \bar{W}^{*j})$. We may thus write domestic government welfare as a function

of the domestic regulatory choices, the foreign regulatory choices and foreign tariffs, and the foreign welfare levels, or

$$W(\mathbf{r}, \bar{p}(\mathbf{r}, \mathbf{r}^{*j}, \tau^{*j}, \bar{W}^{*j}), \bar{T}(\{\mathbf{r}^{*j}\}, \{\tau^{*j}\}, \{\bar{W}^{*j}\})). \quad (\text{A18})$$

Fixing foreign welfare levels and choosing domestic and foreign regulations and foreign tariffs to maximize domestic welfare given by (A18) then defines a point on the efficiency frontier. The first order conditions that define the efficiency frontier are

$$W_{r_i} + W_p \frac{\partial \bar{p}}{\partial r_i} = 0 \quad \text{for } i = 1, 2, \dots, R, \quad (\text{A19})$$

$$W_p \frac{\partial \bar{p}}{\partial r_i^{*j}} + W_{\bar{T}} \frac{\partial \bar{T}}{\partial r_i^{*j}} = 0 \quad \text{for } i = 1, 2, \dots, R^{*j} \text{ and } j = 1, 2, \text{ and} \quad (\text{A20})$$

$$W_p \frac{\partial \bar{p}}{\partial \tau^{*j}} + W_{\bar{T}} \frac{\partial \bar{T}}{\partial \tau^{*j}} = 0 \quad \text{for } i = 1, 2. \quad (\text{A21})$$

Now consider the efficiency properties of policy choices that satisfy (6.11) and also (6.12). By (A19)- (A21), these policies are efficient if and only if

$$\frac{\partial \bar{T}}{\partial r_i^{*j}} = 0 = \frac{\partial \bar{T}}{\partial \tau^{*j}} \quad \text{for } i = 1, 2, \dots, R^{*j} \text{ and } j = 1, 2. \quad (\text{A22})$$

But by (A17), (A22) is satisfied at policies that satisfy (6.11) and (6.12) if and only if

$$\frac{E_{r_i^{*j}}^{*j}}{M} [\tilde{p}^{wj} - \bar{T}] = 0 = \frac{E_{p^{*j}}^{*j}}{M} [\tilde{p}^{wj} - \bar{T}] \quad \text{for } i = 1, 2, \dots, R^{*j} \text{ and } j = 1, 2. \quad (\text{A23})$$

Hence, by (A23), policy choices that satisfy (6.11) and (6.12) are efficient if and only if the tariffs conform to MFN (so that $\tilde{p}^{wj} - \bar{T}$ for $j = 1, 2$). Further, at policies satisfying (6.11), (A17)-(A23) can be used to show that efficiency requires that these policies satisfy (6.12) as well and abide by MFN. Hence, an international agreement can attain a point on the international efficiency frontier and satisfy (6.11) if and only if it satisfies the MFN rule. QED