Plurilateral Trade Agreements: A Complementary Margin to Preferential Liberalization*

Lasha Chochua† TSU ISET

James Lake‡ University of Tennessee

Gerald Willman§ Uni Bielefeld, IfW Kiel

March 18, 2026

Abstract

We show that plurilateral agreements facilitate global tariff liberalization by creating an MFN-based margin of cooperation that leaves preferential access via preferential trade agreements (PTAs) unchanged. In a model of endogenous trade agreement formation with farsighted governments, PTAs become rigid once exclusion or free-riding incentives bind, constraining further PTA expansion. Plurilateral agreements relax these constraints by allowing countries to liberalize selectively in a differentiated-goods sector without altering existing PTAs. As a result, the stable equilibrium trade network consists of the PTAs that would arise absent plurilaterals, augmented—but not replaced—by plurilateral MFN liberalization. This mechanism provides an explanation for the growing role of sectoral plurilateral agreements within the WTO as preferential liberalization becomes increasingly constrained.

Keywords: Plurilateral Agreements, Preferential Trade Agreements, Global Free Trade, WTO.

JEL Classification Numbers: F12, F13, F15, F18.

*For useful comments and suggestions, we thank seminar and conference participants at Bologna, Kiel-Aarhus workshop 2023, Göttingen workhsop 2024, APTS 2024 Bangkok, ETSG 2024 Athens, Griep 2024 Istanbul, and the 2025 EEA Congress. Lasha Chochua and Gerald Willmann gratefully acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG) under Grant WI 2038/6-1.

†E-mail: lasha.chochua@iset.ge

‡E-mail: jlake7@utk.edu

§E-mail: gwillmann@wiwi.uni-bielefeld.de

1 Introduction

The last successful round of comprehensive multilateral trade liberalization concluded with the Uruguay Round in 1994. In the three decades since, repeated efforts to advance global trade negotiations under the World Trade Organization (WTO) have stalled, reflecting deep heterogeneity in national interests, rising geopolitical tensions, and the increasing difficulty of achieving consensus among a large and diverse membership. While the multilateral trading system has proven durable in preventing large-scale reversals toward protectionism and retaining its core institutional functions, the WTO has experienced prolonged gridlock in concluding new rounds of tariff liberalization and rulemaking (Bagwell and Staiger, 2002; Freund and Ornelas, 2010; Wolfe, 2015; Bown and Crowley, 2016). As a result, a central question in international economics is no longer whether multilateral liberalization has slowed, but how trade cooperation evolves once comprehensive global agreements become politically unattainable.

The dominant institutional response to multilateral gridlock has been the rapid proliferation of preferential trade agreements (PTAs). Bilateral and regional agreements now govern a substantial share of world trade and have become the primary vehicle for negotiated liberalization (Bown and Crowley, 2016; Lake, 2026). A large literature studies the formation and consequences of PTAs, with a key theme throughout the classic and more recent literature being the incentives of countries to include or exclude partners from preferential access.[1] By bundling liberalization across sectors and relying on discrimination, PTAs offer a clear and tractable mechanism for reciprocity and help governments internalize gains from cooperation across industries. In this sense, PTAs can be understood as a deliberately rigid institutional form. By linking sector-level tariff cuts within a single agreement, they limit fragmentation and mitigate incentives to cherry-pick sectors for discriminatory liberalization (Davis, 2004; Hoekman and Mavroidis, 2015a; Maggi, 2016).

At the same time, this rigidity implies that preferential liberalization is inherently limited once the scope for the mutually beneficial exchange of reciprocal preferential access is constrained. Dominant countries may prefer to exclude weaker partners to preserve preferential access, while others may choose to free ride on the liberalization efforts of their trading partners. As a result, PTAs can function as a straightjacket on further liberal-

[1] Classic references include Krugman (1991), Baldwin (1995), Krishna (1998), Bagwell and Staiger (1999), and Ornelas (2005a,b,c, 2008), who often addressed the question of whether PTAs act as building or stumbling blocks on the way to global free trade. Later references include Saggi and Yildiz (2010), Saggi et al. (2013), Facchini et al. (2012, 2021), Lake and Yildiz (2016), Lake and Roy (2017), Lake (2017, 2019), Lake et al. (2020) and Berens et al. (2021). For surveys, see Freund and Ornelas (2010), Maggi (2014), and Krishna and Lake (2019).

ization: effective as an initial response to multilateral deadlock, but increasingly binding as distributional incentives prevent further preferential expansion (Saggi and Yildiz, 2010; Saggi et al., 2013; Missios et al., 2016; Lake and Yildiz, 2016; Berens et al., 2021).[2]

Alongside the expansion of PTAs, a distinct and comparatively under-analyzed institutional form has emerged: plurilateral trade agreements that liberalize trade on a non-discriminatory MFN basis within specific sectors.[3] Unlike PTAs, plurilateral agreements do not grant preferential access to members at the expense of outsiders. Instead, participating countries commit to tariff reductions that apply equally to members and non-members, preserving the MFN principle while relaxing the unanimity constraint that has hindered WTO-wide negotiations. In this sense, plurilateral agreements occupy an intermediate institutional position between multilateralism and discriminatory regionalism, allowing subsets of countries to advance liberalization without dismantling existing preferential arrangements (Hoekman and Mavroidis, 2015b; Lewis, 2025).

Plurilateral agreements are not a theoretical curiosity. The WTO Information Technology Agreement (ITA), for example, eliminates MFN tariffs on hundreds of information and communications technology products and covers the vast majority of global trade in those goods (Mann and Liu, 2009; Tang and Lascari, 2017; Gnutzmann-Mkrtchyan and Henn, 2018; Chochua and Iodice, 2025). More recent initiatives—including negotiations over an Environmental Goods Agreement and a range of WTO Joint Statement Initiatives covering e-commerce, services domestic regulation, and investment facilitation—underscore the growing reliance on plurilateral approaches as a pragmatic response to stalled multilateralism (Hoekman and Sabel, 2019, 2021; Wolff, 2023). These agreements already shape global trade flows and are likely to play an increasingly important role going forward.[4]

Scholars in international trade law and international relations increasingly view plurilateral agreements as a form of institutional "variable geometry" within the WTO system. In this view, plurilaterals reconcile non-discrimination with political feasibility in heterogeneous memberships and serve as club-based solutions to collective action problems (Hoekman and Mavroidis, 2015a, 2017). From an international relations perspective, plurilateral

[2]Empirically, many major trading relationships among large economies already operate near this boundary, where PTA formation seems politically constrained despite apparent scope for mutual gains from liberalization.

[3]Plurilateral agreements have also been referred to as "critical mass" agreements (Winslett, 2018), as they become truly effective only if a critical mass of countries participate.

[4]WTO/GATT plurilateral agreements pre-date the ITA. The Agreement on Trade in Civil Aircraft entered into force in 1980 and currently has over 50 member countries that have eliminated their MFN tariffs on non-military aircraft and their parts and components. The Agreement on Trade in Pharmaceutical Products entered into force in 1994 and currently has over 25 member countries that have eliminated their MFN tariffs on pharmaceutical goods and their chemical inputs.

agreements allow coalitions of willing countries to internalize gains from cooperation while preserving openness to future participation (Abbott and Snidal, 2000; Keohane and Victor, 2011). Despite this growing recognition outside economics, formal economic analysis of plurilateral agreements remains limited, particularly with respect to how MFN-based sectoral liberalization interacts with PTAs and reshapes countries' incentives to cooperate.

Our paper addresses this gap by developing a model of endogenous trade agreement formation in which countries choose between PTAs and plurilateral MFN-based liberalization in specific sectors. Our three-country model has a traditional three-good competing-exporters framework that has served as the workhorse model in the literature and allows comparability with prior literature. We augment this framework by adding a differentiated-goods sector where plurilateral liberalization may play out. This combination allows PTAs to generate exclusion and free-riding incentives depending on the structure of supply side asymmetries across countries as in previous work, and also allows plurilateral agreements to generate MFN-based gains that depend on market size and demand elasticities. We consider a comprehensive set of global trade networks including no agreements (i.e. MFN-constrained unilateral policy setting), a full set of PTA regimes, plurilaterals, and combinations of PTAs and plurilaterals.[5] Based on the welfare levels under these trade networks, we find the equilibrium trade network using a farsighted stability concept, the largest consistent set proposed by Chwe (1994). This enables countries to anticipate how initial deviations may trigger subsequent changes in the trade agreement network.

The analysis yields three central insights. First, plurilateral agreements do not undermine or replace existing PTAs. Across the environments we consider, the introduction of plurilateral liberalization does not eliminate PTAs from the set of stable trade agreement networks. Instead, plurilaterals coexist with PTAs, expanding liberalization along MFN-based sectoral margins while leaving preferential access unchanged. This result helps allay concerns—frequently raised in policy discussions—that plurilateral agreements may destabilize the existing trade architecture (Hoekman and Mavroidis, 2015b; Wolff, 2023).

Second, plurilateral agreements relax binding constraints on further liberalization once preferential liberalization has been exhausted. In environments with asymmetric countries—where exclusion or free-riding incentives prevent the expansion of PTAs—plurilateral agreements enable additional global tariff reductions in sectors of mutual interest. In equilib-

[5]From a modeling perspective, the PTA literature has focused on three dimensions: (i) the number of countries (typically three-country or N-country models), (ii) the economic characteristics of countries (which shapes PTA formation incentives), and (iii) the institutional design of trade agreements (e.g. FTAs vs CUs vs multilateral agreements). In accordance with the research question, papers typically do not treat each dimension in-depth. Our research question focuses on the institutional design of trade agreements and hence we naturally focus on dimensions (ii) and (iii).

rium, the stable trade agreement network corresponds to the PTA-based outcome that would arise absent plurilaterals, augmented by MFN-based plurilateral liberalization. Plurilaterals thus operate as an institutional overlay that extends liberalization beyond what preferential agreements alone can achieve.

Third, this mechanism provides a dynamic interpretation of recent developments in the global trading system. PTAs serve as an effective initial response to multilateral deadlock, but their rigidity eventually becomes binding. Plurilateral agreements emerge precisely at this stage, offering the institutional flexibility required to sustain further cooperation without unraveling existing PTAs. Viewed through this lens, the increasing prominence of plurilateral agreements reflects sequential institutional adaptation rather than fragmentation of the trading system (Hoekman and Mavroidis, 2015b; Hoekman and Sabel, 2019, 2021). In this sense, plurilateral agreements emerge endogenously as a response to the institutional limits of preferential liberalization.

The paper is organized as follows. Section 2 introduces the trade model and characterizes optimal tariffs across different trade agreement networks. Section 3 presents the trade agreement formation game and the stability concept. Section 4 analyzes equilibrium trade agreement networks with symmetric countries. Sections 5 and 6 study asymmetric environments with, respectively, exclusion and free-riding incentives. Section 7 concludes.

2 Model

2.1 Underlying Trade Model

Overview. Our underlying trade model has three countries $z = a, b, c$ and four non-numéraire goods $Z = A, B, C, D$. Where necessary, we denote countries (and goods) generically by i, j, k (I, J, K). The numéraire good 0 is produced one-to-one using labor and thus pins down unit wages on the supply side and also absorbs all income effects on the demand side.

Goods $Z = A, B, C$ are homogeneous goods subject to comparative advantage forces and perfect competition as in the traditional competing exporters model.[6] Country $z = i$ has zero endowment of good $Z = I$ but an endowment $e_z > 0$ of the two goods $Z \neq I$. Thus, country $z = i$ imports good $Z = I$ and exports goods $Z \neq I$. We augment this

[6]While we use the competing exporters model for comparability as it has been the workhorse model in this literature, our approach accommodates other underlying trade models that have optimal tariff solutions (if only numerically). Indeed, the equilibrium trade agreement network ultimately depends on country preferences over trade agreement networks. To this end, we consider the presence of free riding and exclusion incentives which have featured prominently in the recent PTA literature.

traditional competing exporters setup with good $Z = D$ which represents a monopolistically competitive differentiated goods sector *à la* Armington that operates under constant marginal cost normalized to 1.[7] Each country z has a set of varieties Ω_z that contains N elements.

The coexistence of homogeneous and differentiated goods sectors allows PTAs and plurilateral agreements to operate along distinct margins. PTAs affect tariffs through discrimination, consistent with the PTA literature in which sectoral comparative advantage generates the gains from reciprocal tariff reductions and gives rise to exclusion and free-riding incentives. In contrast, the selectively chosen, single-sector nature of plurilateral agreements makes them more naturally suited to a differentiated goods sector, where all countries produce and consume distinct varieties of the same good and liberalization operates on an MFN basis. This sectoral distinction is therefore essential for capturing the different economic mechanisms underlying preferential and plurilateral liberalization.[8]

Preferences. Each country i has a representative consumer with quasi-linear preferences that are linear in the numéraire good and additively separable across non-numéraire goods. Subutility is quadratic for homogeneous goods A, B, C and for the *composite* differentiated good D. This implies linear demand at the sector level for non-numéraire goods:

$$x_{iI} = \alpha - p_{iI} \quad \text{for } I = A, B, C, D \tag{1}$$

where α is the choke price that indicates the size of the market, and p_{iI} is the (tariff-inclusive) price paid by country i consumers for good I.

Preferences are nested for the differentiated good D, such that varieties within the composite differentiated good are aggregated via a CES index with elasticity of substitution $\sigma > 1$. The CES aggregate and the implied variety-level demand are

$$x_{iD} = \left(\sum_{j=a,b,c} \sum_{n \in \Omega_j} x_{nji}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{2}$$

$$x_{nji} = x_{iD} \left(\frac{p_{nji}}{p_{iD}} \right)^{-\sigma} \tag{3}$$

where x_{nji} and p_{nji} denote consumption and price of variety n from country j in country i, and p_{iD} is the CES price index (see equation (8) below) for the composite differentiated

[7] Symmetry in the differentiated good sector is a standard assumption and keeps the model tractable.

[8] We abstract from endogenous firm entry to focus on governments' tariff-setting incentives; allowing entry would not overturn the MFN-based gains from plurilateral liberalization emphasized here.

good in country i. This structure preserves linear demand for the composite differentiated good while allowing within-sector substitution across varieties, which is central for MFN-based tariff changes under plurilateral agreements.

Equilibrium in Homogeneous Goods Markets. International market clearing for non-numéraire good I equates the import demand of country i (i.e. x_{iI}) and the export supply from countries $z \neq i$ (i.e. $e_z - x_{zI} = e_z - \alpha + p_{zI}$). Given additive (i.e. specific) tariffs drive a wedge between the tariff-inclusive price p_{iI} paid by country i consumers and the tariff-exclusive price p_{jI} of good I received by country j exporters,

$$p_{iI} = p_{jI} + t_{ij} \text{ for } j \neq i \tag{4}$$

market clearing then yields the following equilibrium prices:

$$p_{iI} = \alpha - \frac{1}{3}\sum_{z \neq i} e_z + \frac{1}{3}\sum_{z \neq i} t_{iz} \text{ for } z = i. \tag{5}$$

The competing exporters structure implies equilibrium prices depend only on the importing country's tariffs and the endowments of the two exporting countries.

Equilibrium in Differentiated Goods Sector. Each firm n in country j produces a single variety recognizing the variety-level demand in country i given by equation (3) and that tariffs drive a wedge between domestic and imported varieties,

$$p_{nji} = p_{nii}\tau_{ij} \text{ for } j \neq i \text{ and } \tau_{ij} \geq 1 \tag{6}$$

where $\tau_{ij} = 1$ ($\tau_{ij} > 1$) represents a zero (strictly positive) advalorem tariff. Maximizing profits delivers the standard markup rule for monopolistically competitive firms

$$p_{njj} = \frac{\sigma}{\sigma - 1}c \quad \text{and} \quad p_{nji} = \frac{\sigma}{\sigma - 1}c\tau_{ij} \tag{7}$$

and the CES price index in country i

$$p_{iD} = \left(N \sum_{j=a,b,c} p_{nji}^{1-\sigma}\right)^{\frac{1}{1-\sigma}} = \frac{\sigma}{\sigma - 1}c\tau_i. \tag{8}$$

Here, $\tau_i \equiv \left(N \sum_{j=a,b,c} \tau_{ij}^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$ represents the inward multilateral resistance term, which summarizes the average level of tariffs that country i imposes on all foreign varieties in the differentiated goods sector. This term plays a central role in determining how MFN-based tariff changes under plurilateral agreements affect market access and welfare.

Given this price index, variety-level demand and composite differentiated good demand from equations (2) and (3) become

$$x_{iD} = \alpha - \frac{\sigma}{\sigma - 1} c\tau_i \tag{9}$$

$$x_{nji} = \left(\alpha - \frac{\sigma}{\sigma - 1} c\tau_i\right) \left(\frac{\tau_{ij}}{\tau_i}\right)^{-\sigma}. \tag{10}$$

Welfare. The quasi-linearity and additive separability of the utility function simplify the welfare comparison across alternative trade networks. In particular, national welfare for country z, denoted W_z, is the sum of consumer surplus and producer surplus for non-numéraire goods and tariff revenue. Producer surplus for the traditional non-numéraire goods $Z = A, B, C$ is the value of endowments whereas producer surplus in the differentiated goods sector consists of firm profits.[9]

2.2 Trade Agreement Networks

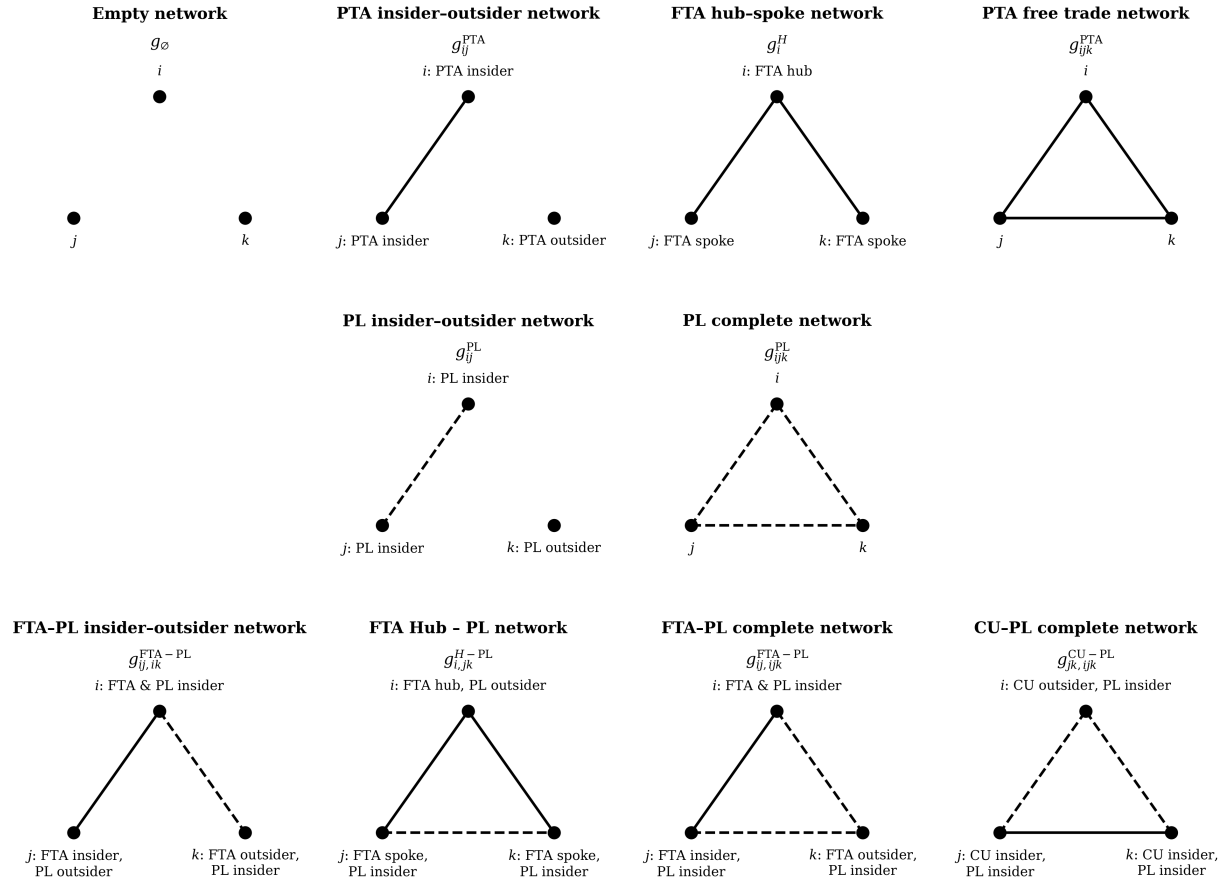
Before deriving optimal tariffs and welfare, we first describe the different trade networks and the positions that countries can occupy in these trade networks. Figure 1 depicts the set of trade agreement networks $\mathbb{G}$. These networks summarize the feasible institutional configurations through which countries may liberalize trade via preferential and plurilateral agreements.

The first row illustrates networks that only contain PTAs where, by definition, PTA members impose zero tariffs on each other on all goods. The empty network $g_\emptyset$ contains no agreements. In PTA insider-outsider networks g_{ij}^{PTA}, countries i and j have a PTA that is either an FTA (g_{ij}^{FTA}) or a CU (g_{ij}^{CU}). FTAs and CUs differ because, unlike FTA members, CU members impose a common external tariff on non-members. There are six PTA insider-outsider networks given three country pairs and two types of PTAs. In FTA hub-spoke networks g_i^H, country i is the "hub" that has FTAs with each "spoke" country j and k. Each country can be the hub, so there are three such networks.[10] A PTA free

[9] Appendix B presents closed-form expressions for consumer surplus, producer surplus and tariff revenue.

[10] There are no CU hub-spoke networks because the common external tariff requirement of a CU implies countries cannot form "overlapping" CUs. The absence of the common external tariff requirement in FTAs allows FTA hub-spoke networks.

trade network g_{ijk}^{PTA} is either three bilateral FTAs (g_{ijk}^{FTA}) or a three-country CU (g_{ijk}^{CU}).[11] Thus, there are two PTA free trade networks.



Notes: A solid line represents a PTA link, a dashed line represents a plurilateral (PL) link between pairs of countries, and the absence of any line indicates that there is neither a PTA nor a PL link.

Figure 1: Trade Networks and Positions

The second row illustrates networks that only contain plurilateral agreements. By definition, members of a plurilateral agreement reduce tariffs in the differentiated goods sector on an MFN basis. In plurilateral insider-outsider networks g_{ij}^{PL}, countries i and j have a plurilateral agreement. There are three such networks since there are three country pairs. We assume "open" plurilateral agreements: the outsider country k can unilaterally join a plurilateral agreement between the insider countries i and j. Given the MFN tariffs in a plurilateral agreement, the world would then move from the plurilateral insider-outsider

[11] Given the zero internal tariff requirement of an FTA, there is no distinction between the tariffs of three bilateral FTAs or a three-country FTA. Thus, we can analytically view a three-country FTA as three bilateral FTAs without loss of generality.

network g_{ij}^{PL} to the plurilateral complete network g_{ijk}^{PL}. The plurilateral complete network is not equivalent to global free trade because it does not necessarily involve zero tariffs and does not cover homogeneous goods.

The third row illustrates networks that contain PTAs and plurilateral agreements. In FTA-PL insider-outsider networks $g_{ij,ik}^{FTA-PL}$, countries i and j have an FTA and countries i and k have a plurilateral agreement.[12] There are six such networks given three country pairs and two types of agreements (FTAs and plurilaterals). In hub-plurilateral networks $g_{i,jk}^{H-PL}$, country i is the hub and has FTAs with the spoke countries j and k while countries j and k have a plurilateral agreement. There are three such networks. Lastly, in the PTA-PL complete network $g_{ij,ijk}^{PTA}$, countries i, j and k have a plurilateral agreement while countries i and j also have either an FTA ($g_{ij,ijk}^{FTA-PL}$) or a CU ($g_{ij,ijk}^{CU-PL}$). There are six such networks.

Overall, Figure 1 depicts 31 possible trade agreement networks. While this set is large, imposing symmetry or limited asymmetry across countries, together with the notion of farsighted stability, sharply restricts the set of candidate stable trade networks and yields a tractable equilibrium characterization.

2.3 Tariffs and Welfare

2.3.1 Homogeneous Goods Sector

Analytical expressions for optimal tariffs and welfare in the homogeneous goods sector are well-known from the prior PTA literature (Saggi and Yildiz, 2010; Saggi et al., 2013; Missios et al., 2016; Lake, 2017; Berens et al., 2021).[13] Due to the reallocation of imports away from the PTA non-member and towards the fellow PTA member following PTA formation, PTA members have an incentive to lower their external tariff on the PTA non-member. That is, PTA members practice tariff complementarity. That said, CU members impose higher external tariffs because they coordinate external tariffs and hence internalize the associated externality that tariff complementarity imposes on each member. In contrast, due to the competing exporters structure whereby each country imports a different good, PTA non-members do not change their tariffs following PTA formation by other countries.

In terms of welfare with symmetric countries, PTA members benefit from the mutual exchange of preferential access but the PTA non-member suffers from the associated discrimination in the PTA member markets. When countries are asymmetric in terms of their

[12] Allowing countries i and j to have a PTA and a plurilateral is redundant. The zero internal tariff requirement of PTAs imply zero bilateral tariffs among members on all goods. And PTA members can choose any tariff on the non-member in the absence of a plurilateral agreement. Moreover, the common external tariff requirement rules out CU-PL insider outsider networks $g_{ij,ik}^{CU-PL}$.

[13] See Appendix Table A.1 for the optimal tariffs in the homogeneous goods sector.

endowment, countries with smaller endowments are more attractive PTA partners because, given symmetric demand across countries, such countries are both "large importers" and "small exporters".[14] The former increases the gain in producer surplus from exporting to the PTA member and the latter reduces the loss in producer surplus from stronger import competition. This naturally leads to "exclusion incentives" and "free riding incentives". With two small countries and one large country, the small countries want to exclude the large country from expansion of their bilateral PTA into a trilateral PTA. In contrast, with one small country and two large countries, the small country prefers to free ride on PTA formation by the large countries rather than form their own bilateral PTA with a large country.

In our subsequent analysis, we therefore consider three cases: symmetric countries where all countries benefit from PTA formation, and the asymmetric cases with exclusion and free riding incentives. Ensuring non-negative exports across all trade agreement networks given the optimal tariffs presented in Appendix Table A.1 requires $e_l/e_s \leq 5/3$. To maximize the constraints imposed on PTA liberalization by the exclusion and free riding incentives, we impose $e_l/e_s = 5/3$ when countries are asymmetric.[15]

2.3.2 Tariffs and Welfare in the Differentiated Goods Sector

Unlike the homogeneous goods sector, welfare in the differentiated goods sector is typically highly non-linear and precludes closed-form solutions for optimal tariffs and welfare. An important exception is the empirically plausible case of $\sigma = 2$, which allows for transparent analytical expressions. We therefore use this case to summarize optimal external tariffs in both the differentiated and homogeneous goods sector across trade networks in Table 1. Importantly, the economic forces driving optimal tariffs and welfare are not specific to $\sigma = 2$ and extend to more general values of σ. Accordingly, we do not restrict the subsequent analysis of stable trade networks to this special case.

Optimal external tariffs are increasing in N and α in all trade networks. An increase in N or α raises import volumes and strengthens terms-of-trade motivations, leading governments to set higher external tariffs. To begin, the optimal (MFN) tariff in the absence of any agreements is

$$\tau_{ij}\left(g_{\emptyset}\right) = \frac{2(-2+\sqrt{3+N^2\alpha^2})}{-1+N\alpha} \text{ for } j \neq i. \tag{11}$$

[14] We assume countries have sufficient endowments of the numéraire good to balance bilateral trade.

[15] Non-negative prices across all trade agreement networks given the optimal tariffs presented in Appendix Table A.1 requires $\alpha \gtrsim 1.4$. To ensure strictly positive positive prices, we impose $\alpha \geq 1.5$.

Network	Countries	Homogeneous	Differentiated
$g_{\emptyset}$	All	$t_{ij}(g_{\emptyset})$	$\tau_{ij}(g_{\emptyset})$
g_{ij}^{FTA}	Insiders (i,j)	$t_{ik}(g_{ij}^{FTA})$	$\tau_{ik}(g_{ij}^{FTA})$
	Outsider (k)	$t_{ki}(g_{\emptyset})$	$\tau_{ki}(g_{\emptyset})$
g_{ij}^{CU}	Insiders (i,j)	$t_{ik}(g_{ij}^{CU})$	$\tau_{ik}(g_{\emptyset})$
	Outsider (k)	$t_{ki}(g_{\emptyset})$	$\tau_{ki}(g_{\emptyset})$
g_{i}^{H}	Spokes (j,k)	$t_{jk}(g_{ij}^{FTA})$	$\tau_{jk}(g_{ij}^{FTA})$
g_{ij}^{PL}	Insiders (i,j)	$t_{ik}(g_{\emptyset})$	$\tau_{ik}(g_{ij}^{PL})$
	Outsider (k)	$t_{ki}(g_{\emptyset})$	$\tau_{ki}(g_{\emptyset})$
g_{ijk}^{PL}	All	$t_{ij}(g_{\emptyset})$	1^{*}
$g_{ij,ik}^{FTA-PL}$	FTA-PL insider (i)	$t_{ik}(g_{ij}^{FTA})$	1^{*}
	FTA insider (j)	$t_{jk}(g_{ij}^{FTA})$	$\tau_{jk}(g_{ij}^{FTA})$
	PL insider (k)	$t_{kj}(g_{\emptyset})$	$\tau_{kj}(g_{ik}^{PL})$
$g_{i,jk}^{H-PL}$	Spokes (j,k)	$t_{jk}(g_{ij}^{FTA})$	1^{*}
$g_{ij,ijk}^{PTA-PL}$	PTA insiders (i,j)	$t_{ik}(g_{ij}^{PTA})$	1^{*}
	PTA outsider (k)	$t_{ki}(g_{ij}^{PTA})$	1^{*}

Table 1: Network-dependent Optimal External Tariffs

Notes: See Figure 1 and text for more details. $\tau \geq 1$ implies $\tau = 1$ is a "zero" tariff. * indicates external tariff $\tau = 1$ imposed because otherwise import subsidies optimal. Plurilateral members impose external tariff on all countries. PTA members impose zero internal tariffs.

In PTA insider-outsider networks g_{ij}^{PTA}, the PTA outsider continues imposing its MFN tariff: $\tau_{ki}\left(g_{ij}^{PTA}\right) = \tau_{ki}\left(g_{\emptyset}\right)$. This follows from a constant marginal cost because, in turn, the tariffs imposed by PTA members do not affect the incentives of firms to export to the PTA outsider. In contrast, the compositional shift in imports towards its PTA member partner implies PTA members have an incentive to adjust their optimal tariff on the PTA outsider k. FTA insiders set their external tariffs independently

$$\tau_{ik}\left(g_{ij}^{FTA}\right) = \frac{-10 + 5N\alpha + \sqrt{60 + 11N\alpha(-12 + 11N\alpha)}}{2(-5 + 6N\alpha)} \tag{12}$$

and, as with homogeneous good tariffs, practice tariff complementarity by lowering their tariff on the FTA outsider (i.e. $\tau_{ik}\left(g_{ij}^{FTA}\right) < \tau_{ij}\left(g_{\emptyset}\right)$). Because CU insiders coordinate their external tariff, they set an external tariff of

$$\tau_{ik}\left(g_{ij}^{CU}\right) = \frac{-5 + 3N\alpha + \sqrt{21 + 5N\alpha(-6 + 5N\alpha)}}{-2 + 4N\alpha}, \tag{13}$$

which exceeds the external tariff of FTA insiders (i.e. $\tau_{ik}\left(g_{ij}^{CU}\right) > \tau_{ik}\left(g_{ij}^{FTA}\right)$). Indeed, the optimal CU tariff exceeds the optimal MFN tariff (i.e. $\tau_{ik}\left(g_{ij}^{CU}\right) > \tau_{ik}\left(g_{\emptyset}\right)$). Because this would violate the GATT Article XXIV requirement that PTA members do not raise tariffs on non-members, we impose $\tau_{ik}\left(g_{ij}^{CU}\right) = \tau_{ik}\left(g_{\emptyset}\right)$.

When the FTA insider-outsider network g_{ij}^{FTA} transitions to the hub-spoke network g_i^H, the hub sets zero tariffs due to its bilateral FTAs with each spoke country. Based on the logic underlying the optimal tariffs at the FTA insider-outsider network g_{ij}^{FTA}, each spoke country $z, z' = j, k$ sets the optimal tariff of an FTA insider: $\tau_{zz'}\left(g_i^H\right) = \tau_{jk}\left(g_{ij}^{FTA}\right)$. Upon transition to the PTA free trade network g_{ijk}^{PTA}, all countries set zero tariffs.

In networks with plurilateral agreements, plurilateral members choose their MFN tariffs to maximize their joint welfare. In the plurilateral insider-outsider network, their optimal tariff is

$$\tau_{iz}\left(g_{ij}^{PL}\right) = \frac{-14 - N\alpha + \sqrt{148 + 7N\alpha(4 + 7N\alpha)}}{-4 + 4N\alpha} \text{ for } z = j, k. \tag{14}$$

Again, constant marginal cost implies the optimal external tariff for the plurilateral outsider remains unchanged at its MFN tariff: $\tau_{ki}\left(g_{ij}^{PL}\right) = \tau_{kz}\left(g_{\emptyset}\right)$ for $z = i, j$. In the PL complete network g_{ijk}^{PL}, the optimal MFN tariff entails import subsidies. Although WTO rules do not expressly prohibit import subsidies, such policies are rarely observed and widely viewed as politically infeasible. We therefore impose a non-negativity constraint on tariffs. As such, $\tau_{ij}\left(g_{ijk}^{PL}\right) = 1$.

In networks with PTAs and plurilateral networks, we combine the above insights. In an FTA-PL insider–outsider $g_{ij,ik}^{FTA-PL}$ network, i is an insider in both agreements, j is an FTA (plurilateral) insider (outsider), and k is an FTA (plurilateral) outsider (insider). The FTA insider j keeps their FTA insider optimal tariff: $\tau_{jk}\left(g_{ij,ik}^{FTA-PL}\right) = \tau_{jk}\left(g_{ij}^{FTA}\right)$. The PL insider keeps their optimal tariff from the PL insider-outsider network: $\tau_{kj}\left(g_{ij,ik}^{FTA-PL}\right) = \tau_{kj}\left(g_{ik}^{PL}\right)$. For the insider in both agreements, i, they would optimally impose import subsidies and so we constrain this to a zero tariff: $\tau_{ik}\left(g_{ij,ik}^{FTA-PL}\right) = 1$. Similarly, spoke countries in the hub-PL network g_i^{H-PL} and the PL complete network g_{ijk}^{PL} would optimally impose import subsidies and hence we constrain these to zero tariffs: $\tau_{jk}\left(g_{i,jk}^{H-PL}\right) = \tau_{ij}\left(g_{ijk}^{PL}\right) = 1$

3 Trade Agreement Formation Game

3.1 Overview

Our trade agreement formation game has four elements. First, $\mathbb{C}$ is the set of players, consisting of three countries generically denoted i, j and k. Second, Figure 1 depicts the

set of possible trade networks $\mathbb{G}$ and allows us to describe how a coalition $S \subseteq \mathbb{C}$ can deviate from a network g to another network g'. It can do so if and only if (i) S contains at least one country in each agreement severed when moving from g to g' and (ii) S contains all countries in each new agreement when moving from g to g'. We denote such a feasible transition by $g \rightarrow_S g'$. Hereafter, we simply refer to $g \rightarrow_S g'$ as a deviation with the implicit understanding that it is indeed a feasible deviation. Third, countries have preferences over trade agreement networks. $g' \succ_S g$ ($g' \succeq_S g$) denotes that every country $i \in S$ receives a strictly (weakly) higher payoff under g' rather than g.[16]

The fourth element of our trade agreement formation game is the solution concept. Trade agreements are durable institutional arrangements, and governments understand that changing the existing agreement structure alters the incentives and feasible responses of other countries. Because the central question of this paper concerns how PTAs and plurilateral agreements interact, it is essential to allow countries to reason introspectively about how others would respond to a contemplated deviation from the status quo trade network. Importantly, PTAs and plurilateral agreements liberalize trade along fundamentally different margins: PTAs rely on bilateral, discriminatory liberalization across all sectors, while plurilateral agreements entail open-access MFN liberalization within a single sector. These institutional differences naturally imply that different coalitions of countries may become relevant as the trade agreement network evolves.

Much of the recent game-theoretical literature on PTA formation relies on strategy-based solution concepts such as Coalition-Proof Nash Equilibrium (CPNE) or Markov Perfect Equilibrium (MPE). These approaches are well suited to environments in which a single type of agreement is negotiated through a specified bargaining protocol. However, CPNE restricts how coalitions may evolve following an initial deviation, as subsequent deviations must be undertaken by subsets of the initially deviating coalition. In environments where discriminatory and MFN-based liberalization coexist, such restrictions are conceptually ill-suited: the fundamentally different nature of PTAs and plurilateral agreements implies that coalitions may naturally reconfigure along a sequence of deviations as countries respond to different margins of liberalization. These issues arise naturally in our setting. For example, following a bilateral deviation to form an FTA, each FTA partner will naturally contemplate whether its partner may subsequently form a plurilateral agreement with the FTA outsider. Likewise, following a bilateral deviation to form a plurilateral agreement, each member will naturally contemplate whether the third country may subsequently join the open-access agreement.

[16] Naturally, the singleton coalition $S = \{i\}$ is a possible coalition. Further, $g \succ_S g'$ does not hold if and only if $g' \succeq_i g$ for at least one $i \in S$.

In addition to these conceptual limitations, strategy-based solution concepts require the specification of detailed strategies and bargaining protocols. This becomes increasingly cumbersome as the set of possible trade agreement networks expands and the space of feasible strategies grows. Because our analysis considers a rich space of trade agreement networks involving both PTAs and plurilateral agreements, imposing a particular bargaining protocol or strategy space would substantially complicate the analysis without adding insight. By contrast, the recursive structure of our stability-based concept allows the set of stable trade agreement networks to be computed algorithmically even when the space of possible networks is large, a feature we exploit in our analysis.

Accordingly, we characterize stability using the Largest Consistent Set (LCS) of Chwe (1994). This concept allows countries to evaluate whether a deviation from the status quo would ultimately be profitable, taking into account the anticipated responses of other countries. In doing so, the LCS captures farsighted, counterfactual reasoning through sequences of hypothetical coalitional deviations, in which any non-empty subset of countries may temporarily coordinate on a policy change, anticipating subsequent responses by others. This approach enables us to analyze how the coexistence of PTAs and plurilateral liberalization shapes the set of stable trade agreement networks, without explicitly modeling strategies, dynamic bargaining, or sequential agreement formation.[17]

3.2 Solution Concept: The Largest Consistent Set

The building block of strategy-based or stability-based game theoretic concepts is some notion of "dominance". For stability concepts, an obvious notion of dominance is *direct dominance* that says a coalition S is immediately better off by deviating from g to g':

Definition 1 *(Direct dominance) $g \in \mathbb{G}$ is directly dominated by $g' \in \mathbb{G}$, denoted by $g < g'$, if there exists a coalition $S \subseteq \mathbb{C}$ such that $g \rightarrow_S g'$ and $g' \succ_S g$.*

However, farsighted notions of dominance require players to compare the current outcome with the eventual outcome, rather than the immediate outcome, set in motion by an initial deviation. This is captured by the notion of *indirect dominance*:

Definition 2 *(Indirect dominance) A network $g \in \mathbb{G}$ is indirectly dominated by network $g' \in \mathbb{G}$, denoted by $g \ll g'$, if there exists a sequence of networks $g_1, ..., g_m \in \mathbb{G}$ with $g_1 = g$ and $g_m = g'$ and a sequence of (non-empty) coalitions $S_1, ..., S_m \subseteq \mathbb{C}$ such that $g_n \rightarrow_{S_n} g_{n+1}$ and $g' \succ_{S_n} g_n$ for all $n = 1, .., m-1$.*

[17] While we provide proofs for the LCS in key regions of the parameter space in Appendix C, we use an algorithmic implementation of the LCS to compute stable trade agreement networks numerically across the full parameter space. See Appendix D for more details.

Using the notion of indirect dominance, we now formally introduce the notion of *farsighted deterrence*. This will allow us to define a consistent set and then ultimately arrive at our solution concept of the largest consistent set.

Definition 3 *(Farsighted deterrence) A deviation $g \to_S g'$ by a coalition S is farsightedly deterred by g'' if $g \succeq_S g''$ and either (i) $g' = g''$ or (ii) $g' \ll g''$.*

That is, the deviation by S from g to g' is farsightedly deterred if either (i) it leaves them immediately worse off (i.e. letting $g' = g''$) or (ii) there exists a sequence of coalitional deviations that lead to g'' and every deviating coalition $S_1, ..., S_m$ along the sequence prefers the ultimate outcome g'' over the status quo at the time of their deviation but the initial deviating coalition S_0 does not (strictly) prefer g'' over g.

This allows us to define a *consistent set* in terms of farsighted deterrence.

Definition 4 *(Consistent set) A set of networks $G \subseteq \mathbb{G}$ is a consistent set if and only if every transition $g \to_S g'$ from any $g \in G$ is farsightedly deterred by some $g'' \in G$.*

Intuitively, a set of networks G is consistent if the networks in G farsightedly deter any deviations from networks in G.

This suggests a natural iterative process to find the *largest* consistent set (LCS).[18] To begin, use the full set of networks $\mathbb{G}$ to find the subset of networks G_1 such that any deviation is farsightedly deterred by some $g'' \in \mathbb{G}$. More formally, G_1 can be described as

$$G_1 \equiv f(\mathbb{G}) = \{g : \forall\ g' \text{ and } S,\ g \to_S g' \text{ is farsightedly deterred by some } g'' \in \mathbb{G}\}. \quad (15)$$

Using Definition 4, $\mathbb{G}$ is the the LCS if and only if $G_1 = \mathbb{G}$. In this case, we have reached a fixed point, and the iterative process stops.

However, the iterative process continues if $G_1 \neq \mathbb{G}$. In this case, use the set of networks $G_1 \equiv f(\mathbb{G})$ to find the subset of networks G_2 such that any deviation can be farsightedly deterred by some $g'' \in G_1$. More formally, G_2 can be described as

$$G_2 \equiv f(G_1) = \{g : \forall\ g' \text{ and } S,\ g \to_S g' \text{ is farsightedly deterred by some } g'' \in G_1\}. \quad (16)$$

Again, G_1 is the LCS if and only if $G_2 = G_1$ and, in this case, the iterative process stops. But, the iterative process continues if $G_2 \neq G_1$ until $f^{j+1}(\mathbb{G}) = f^j(\mathbb{G})$, where f^j denotes that f is iteratively applied j times to $\mathbb{G}$. At that point, G_j is the LCS.

[18] More formally, Chwe (1994) establishes that there exists a unique set $G \subseteq \mathbb{G}$ such that G is consistent and every other consistent set $G' \subseteq \mathbb{G}$ satisfies $G' \subseteq G$. This set G is called the largest consistent set.

This iterative and algorithmic-like structure significantly simplifies the process of finding the LCS in practice. We now illustrate this with an example.

Example 1. Naturally, the LCS rests on the welfare rankings across the set of possible trade networks. For illustration and comparability with prior literature, this example ignores plurilateral networks and defines the set of possible networks $\mathbb{G}$ as those in Figure 1 excluding networks with plurilateral agreements. Moreover, we assume that countries hold the following preferences across trade networks: (i) the only network that delivers a higher payoff than global free trade is being the hub, and (ii) being a PTA insider delivers a higher payoff than being a spoke. That is,

$$g_i^H \succ_i g^{FT} \succ_i g \text{ for } g \notin \{g_i^H, g^{FT}\} \text{ and } g^{FT} \in \{g_{ijk}^{CU}, g_{ijk}^{FTA}\} \tag{17}$$

$$g_{ij}^{PTA} \succ_j g_i^H \text{ for } g_{ij}^{PTA} \in \{g_{ij}^{CU}, g_{ij}^{FTA}\}. \tag{18}$$

This is a typical preference structure in various underlying trade models, including our own in this paper, with symmetric countries.

The first step in finding the LCS is to find the set $G_1 = f(\mathbb{G})$ defined by equation (15). This is the set of networks g such that any deviation from g to g' by a coalition of countries S, denoted $g \to_S g'$, is farsightedly deterred by some network $g'' \in \mathbb{G}$. Naturally, farsighted deterrence requires $g \succeq_S g''$ so that S is ultimately (weakly) worse off under g'' than before it deviated away from g. But it also requires that either (i) $g'' = g'$, so that g'' is reached directly from g by S's initial deviation, or (ii) g'' is reached indirectly via a "farsighted" sequence of deviations, denoted $g' \ll g''$ (i.e. g'' indirectly dominates g'). The "farsighted" aspect of the sequence is that each deviating coalition receives a higher payoff under the final network g'' than the status quo network at the time of deviation.

To rule out a network g from the LCS, we need to find a deviation by a coalition S from g to g' that *cannot* be farsightedly deterred. For any network g other than hub-spoke networks or three-country CUs or FTAs, the three-country deviation to the three-country CU $g' = g_{ijk}^{CU}$ cannot be farsightedly deterred. In this case, each country receives a strictly higher payoff under g' than under g and hence farsighted deterrence requires $g' = g_{ijk}^{CU} \ll g''$. But, equation (17) says the only network g'' that any country prefers over $g' = g_{ijk}^{CU}$ is country i being the hub at $g'' = g_i^H$. As such, the only coalition that will initiate a sequence of deviations away from $g' = g_{ijk}^{CU}$ is $S = i$. However, the only deviation available to i is backing out of the three-country CU and leaving the other countries as CU insiders. And, at that point, equation (18) says the CU insiders will not engage in any subsequent deviations that would make them a spoke. Thus, the initial deviation by the three-country coalition from any $g \notin \left\{g_{ijk}^{CU}, g_{ijk}^{FTA}, g_i^H\right\}$ to $g' = g_{ijk}^{CU}$ cannot be farsightedly

deterred and, hence, $g \notin f(\mathbb{G})$ for any such g.

Similar logic establishes that $g_i^H \notin f(\mathbb{G})$. From any hub–spoke network g_i^H, the coalition of the two spoke countries $S = \{j, k\}$ can jointly deviate to the three-country FTA g_{ijk}^{FTA}, which they strictly prefer to remaining spokes. This deviation cannot be farsightedly deterred. The only network that any country prefers is the hub country i returning to g_i^H. However, country i cannot unilaterally restore g_i^H from g_{ijk}^{FTA}. Any further deviation reestablishing g_i^H would require the participation of at least one spoke country, which would be worse off than as an FTA insider. Consequently, no farsighted sequence of deviations deters the initial deviation from g_i^H and, hence, $g_i^H \notin f(\mathbb{G})$.

We can now establish $G_1 = f(\mathbb{G}) = \{g_{ijk}^{CU}, g_{ijk}^{FTA}\}$. As discussed above, the only network that any country i prefers over $g \in G_1$ is being the hub under g_i^H, but country i cannot deviate itself from $g \in G_1$ to g_i^H.[19] Thus, any deviation by a coalition S from $g \in G_1$ to any $g' \notin G_1$ is farsightedly deterred by the subsequent deviation by a two or three-country coalition S' from g' to some $g'' \in G_1$. That is, each country has the same payoff at $g \in G_1$ before any deviations and at $g'' \in G_1$ after the sequence of coalitional deviations. Moreover, payoff equivalence of g_{ijk}^{CU} and g_{ijk}^{FTA} implies the deviation between these two networks is farsightedly deterred (using (i) in Definition 3). Thus, $G_1 = f(\mathbb{G}) = \{g_{ijk}^{CU}, g_{ijk}^{FTA}\}$.

The next step in the iteration process is to find $G_2 \equiv f(G_1)$, as $G_1 = \{g_{ijk}^{CU}, g_{ijk}^{FTA}\}$ is the LCS if $f(G_1) = G_1$. Indeed, this is the case. Note that the only difference between the definitions of $f(\mathbb{G})$ in equation (15) and $f(G_1)$ in equation (16) is that the final network g'' in a farsighted sequence of coalitional deviations is restricted in equation (16) to $g'' \in G_1$ but can be any $g'' \in \mathbb{G}$ in equation (15). However, we only used $g'' \in G_1$ in the previous paragraph to establish $f(\mathbb{G}) = \{g_{ijk}^{CU}, g_{ijk}^{FTA}\} \equiv G_1$. Thus, the same logic establishes that $f(G_1) = G_1$ for $G_1 = \{g_{ijk}^{CU}, g_{ijk}^{FTA}\}$ and, hence, $G_1 = \{g_{ijk}^{CU}, g_{ijk}^{FTA}\}$ is the LCS.

4 Stable Trade Networks with Symmetric Countries

We begin by analyzing the formation of trade agreement networks when countries are symmetric. This benchmark serves two purposes. First, it isolates the interaction between PTAs and plurilateral agreements in the absence of distributional asymmetries that generate exclusion or free-riding incentives. Second—and more importantly—it establishes that allowing plurilateral agreements does not eliminate PTA outcomes. Instead, plurilaterals expand MFN liberalization while leaving preferential access unchanged. Building on this

[19] We could now use part (i) of Definition 3 to establish any deviation from $g \in G_1$ is farsightedly deterred. However, not doing so will help establish $G_1 = f(G_1)$ easily below.

benchmark, the following sections show that the primary role of plurilateral agreements emerges once preferential liberalization becomes constrained.

Throughout this section, countries are symmetric in their endowments in the homogeneous goods sector and share identical demand parameters in the differentiated goods sector, including the elasticity of substitution σ, the number of firms N, and the demand intercept α. This symmetry allows us to isolate the interaction between PTAs and plurilateral agreements without confounding distributional considerations. To highlight the role of plurilateral liberalization, we compare a constrained institutional environment in which countries may form FTAs and CUs but not plurilateral agreements with an extended environment in which plurilateral agreements are also feasible.

For expositional clarity, we first present results for the analytically tractable case of $\sigma = 2$ in the differentiated goods sector. As shown later, the qualitative structure of the LCS carries over to the general case of $\sigma \in (1, \infty)$. Finally, to assess the role of farsighted reasoning, we compare the LCS-based outcomes with those that would arise under coalition-proof Nash equilibrium, highlighting how anticipatory reasoning affects the scope for plurilateral liberalization.

4.1 The Special Case: $\sigma = 2$

4.1.1 Constrained Institutional Framework: No Plurilaterals

As in Example 1, the welfare rankings across the set of possible trade networks dictate the stable networks described by the LCS. Moreover, the relevant welfare rankings in this section are those from Example 1. First, the only outcome better for any country than global free trade is being the hub because this affords preferential access in both spoke markets. Second, countries prefer being PTA insiders over becoming a spoke because they would lose preferential market access with their PTA insider partner and also face discrimination in the other spoke market. That is,

$$g_i^H \succ_i g^{FT} \succ_i g \quad \text{for } g \notin \{g_i^H, g^{FT}\} \text{ and } g^{FT} \in \{g_{ijk}^{CU}, g_{ijk}^{FTA}\} \tag{19}$$

$$g_{ij}^{PTA} \succ_j g_i^H \quad \text{for } g \in \{g_{ij}^{CU}, g_{ij}^{FTA}\}. \tag{20}$$

Proposition 1 summarizes the LCS.

Proposition 1 *Suppose $\sigma = 2$ and $N \geq 3$. Then, the payoff rankings in equations (19)-20) are satisfied when countries are symmetric. Given these preferences, the LCS consists of the three-country CU g_{ijk}^{CU} and the three-country FTA g_{ijk}^{FTA}.*

This outcome is not surprising: global free trade through PTA formation among all countries is the typical equilibrium prediction in endogenous PTA formation models with symmetric countries. At the same time, it serves as a useful benchmark for our analysis, as it confirms that the LCS solution concept reproduces standard results in settings where PTAs alone are sufficient to deliver global free trade in equilibrium. This benchmark provides a natural point of comparison for assessing how the set of stable trade agreement networks changes when plurilateral agreements are introduced.

The proof of Proposition 1 was given in Example 1. Intuitively, it consisted of three parts. First, the LCS can only contain hub-spoke networks g_i^H or a three-country PTA g_{ijk}^{PTA} because the the three-country deviation from any other network to g_{ijk}^{CU} cannot be farsightedly deterred. Equation (19) says the only network that any country possibly prefers is being the hub, but any deviation by the potential hub through backing out of the CU leaves the other countries as the CU insiders. And, equation (20) says the CU insiders would not engage in subsequent deviations that ultimately leaves them as spokes. Second, the hub-spoke network cannot be in the LCS because similar logic implies the deviation by the spokes from g_i^H to g_{ijk}^{FTA} cannot be farsightedly deterred. Third, the three-country PTAs g_{ijk}^{FTA} and g_{ijk}^{CU} constitute the LCS. Given no coalition of countries can directly deviate from g_{ijk}^{CU} or g_{ijk}^{FTA} to being a hub, then any deviation away from g_{ijk}^{CU} or g_{ijk}^{FTA} can be farisghtedly deterred.

4.1.2 Extended Institutional Framework: Plurilaterals

Relative to equations (19)-(20), allowing plurilateral agreements alters the relevant payoff rankings through the introduction of hub-plurilateral networks $g_{i,jk}^{H-PL}$. These impact country preferences in two key ways. The first impact comes from comparing equations (19) and (21). Specifically, the preferential access that i maintains in the traditional markets of the spoke countries after they form a plurilateral means that i still prefers $g_{i,jk}^{H-PL}$ over g_{ijk}^{PTA} even though its preferential market access as the hub is diluted by the plurilateral. The second impact comes from equation (23). On one hand, the spokes in $g_{i,jk}^{H-PL}$ benefit from reciprocal tariff cuts in the differentiated goods sector implemented through their plurilateral agreement (i.e. $g_{i,jk}^{H-PL} \succ_j g_i^H$) and benefit relative to no agreements (i.e. $g_{i,jk}^{H-PL} \succ_j g_\emptyset$). On the other hand, the benefits of plurilateral liberalization are not large

enough to overcome the loss of preferential access as CU insiders (i.e. $g_{ij}^{CU} \succ_{jk} g_{i,jk}^{H-PL}$).

$$g_i^H \succ_i g_{i,jk}^{H-PL} \succ_i g^{FT} \succ_i g \text{ for } g \notin \{g_i^H, g_{i,jk}^{H-PL}, g^{FT}\} \text{ and } g^{FT} \in \{g_{ijk}^{FTA}, g_{ijk}^{CU}\} \tag{21}$$

$$g_{jk}^{PTA} \succ_j g_i^H \text{ for } g_{jk}^{PTA} \in \{g_{jk}^{CU}, g_{jk}^{FTA}\} \tag{22}$$

$$g_{jk}^{CU} \succ_j g_{i,jk}^{H-PL} \succ_j g \text{ for } g \in \{g_i^H, g_\emptyset\} \tag{23}$$

Proposition 2 establishes that plurilaterals emerge in the LCS.

Proposition 2 *Suppose $\sigma = 2$ and $N \geq 3$. Then, the payoff rankings in equations (21)-(23) are satisfied when countries are symmetric. Given these preferences, the LCS consists of the three-country PTAs $g_{ijk}^{CU}, g_{ijk}^{FTA}$ and the hub-plurilateral networks $g_{i,jk}^{H-PL}$, where country i is the hub and the spoke countries j and k are linked by a plurilateral agreement.*

As discussed above, global free trade is the typical equilibrium outcome in endogenous PTA formation models with symmetric countries. Against this benchmark, the emergence of plurilateral agreements in equilibrium is a strong result: even when PTAs alone can deliver global free trade, MFN-based liberalization in selected sectors alters the incentives underlying trade agreement formation. Allowing plurilateral agreements expands the set of stable trade agreement networks relative to the PTA-only benchmark, without eliminating global free trade or PTAs from equilibrium. In symmetric environments—where global free trade is already attainable via PTAs alone—this expansion reflects an additional coordination outcome rather than a failure of liberalization. This benchmark therefore sets the stage for the central policy-relevant question addressed in the subsequent sections: how plurilateral agreements affect PTA formation when global free trade is no longer attainable through PTAs alone.

The intuition underlying Proposition 2 mirrors that of Proposition 1, with one crucial difference. Three-country PTAs remain stable and networks outside this set are still eliminated by farsighted deviations. However, the presence of plurilateral agreements changes the set of stable trade agreement networks by additionally including hub-plurilateral networks $g_{i,jk}^{H-PL}$. Although the spoke countries j and k would prefer to deviate from a hub-plurilateral network to form a three-country FTA, such deviations are farsightedly deterred.

This deterrence arises because the spokes anticipate a sequence of subsequent coalitional deviations that ultimately returns the world to the hub-plurilateral network. For example, starting from g_{ijk}^{FTA}, country i can trigger a sequence of farsighted deviations $g_{ijk}^{FTA} \rightarrow_i g_j^H \rightarrow_{ik} g_\emptyset \rightarrow_{ijk} g_{i,jk}^{H-PL}$ such that each deviating coalition strictly prefers the eventual outcome $g_{i,jk}^{H-PL}$ to the prevailing network. Anticipating this sequence, the initial deviation from the hub-plurilateral network is deterred, and $g_{i,jk}^{H-PL}$ belongs to the LCS.

The key insight is not that plurilateral agreements replace preferential liberalization, but that they reshape the strategic landscape.[20] Even when PTAs alone are sufficient to deliver global free trade, the availability of MFN-based liberalization in a single sector expands the set of stable trade agreement networks. As a result, countries may sustain hybrid trade architectures that combine preferential access in traditional sectors with multilateral-style liberalization in selected industries.

The next sections explore environments with asymmetric countries, where global free trade is not attained in the absence of plurilateral agreements. Before turning to these cases, we briefly compare the LCS-based equilibrium outcomes with those obtained under the more standard solution concept of Coalition-Proof Nash Equilibrium (CPNE).

4.1.3 Comparison with Coalition-Proof Nash Equilibrium

The logic establishing hub-plurilateral networks as part of the LCS highlights an important conceptual distinction between our stability-based approach and the strategy-based solution concepts commonly used in the PTA formation literature, most notably Coalition-Proof Nash Equilibrium (CPNE). While both concepts allow for coalition deviations, they differ in how countries reason about the responses such deviations may trigger. This distinction matters because PTAs and plurilateral agreements liberalize trade along fundamentally different margins, implying that expectations about how members and non-members respond to PTA formation and MFN liberalization are central to the stability of plurilateral agreements.

Under CPNE, a deviation is considered self-enforcing if no subset of the deviating coalition has an incentive to further deviate, holding fixed the actions of countries outside the coalition. As a result, CPNE restricts attention to deviation sequences in which only members of the immediately preceding deviating coalition may respond. In contrast, the LCS allows countries to reason more broadly about how alternative coalitions may subsequently form following an initial deviation, without imposing restrictions on the structure of coalitions along a deviation path.

This distinction is crucial for the emergence of hub-plurilateral networks with symmetric countries. Although such networks belong to the LCS, they are not CPNE outcomes. To see this, consider the hub-plurilateral network $g^{H-PL}_{i,jk}$. Under CPNE, the spoke countries j and k can profitably deviate by forming a three-country FTA, moving the world to g^{FTA}_{ijk}. Because CPNE only allows for subsequent deviations by j and/or k, and no such

[20] In terms of global welfare, the hub-plurilateral network $g^{H-PL}_{i,jk}$ and the global welfare maximizing network g^{PTA}_{ijk} of global free trade are nearly equivalent because the two networks differ only at the margin of preferential access that the hub has in the homogeneous goods sector of each spoke market.

profitable deviations exist, the initial deviation is deemed self-enforcing. Consequently, $g_{i,jk}^{H-PL}$ is ruled out as a CPNE outcome.

By contrast, $g_{i,jk}^{H-PL}$ remains stable under the LCS because countries anticipate a broader sequence of farsighted responses. Starting from g_{ijk}^{FTA}, country i can initiate a sequence of coalitional deviations $g_{ijk}^{FTA} \rightarrow_i g_j^H \rightarrow_{ik} g_\emptyset \rightarrow_{ijk} g_{i,jk}^{H-PL}$ such that each deviating coalition strictly prefers the eventual outcome $g_{i,jk}^{H-PL}$ to the prevailing network at the point of deviation. Anticipating this sequence, the initial deviation by j and k is farsightedly deterred, and the hub-plurilateral network belongs to the LCS.

This comparison shows that the emergence of plurilateral agreements in our framework is not an artifact of equilibrium selection. Rather, it reflects the role of farsighted reasoning in sustaining trade agreements as durable institutions in environments where different forms of liberalization coexist.

4.2 Moving Beyond the Special Case

We now extend the analysis to the general case with $\sigma > 1$. When plurilateral agreements are not available, the result from Section 4.1.1 continues to hold: the LCS consists exclusively of three-country PTAs. The reason is that the payoff rankings in equations (19)-(20) remain satisfied for all values of σ and α when countries are symmetric. Although each country would prefer to be the hub relative to global free trade, spokes strictly prefer being PTA insiders to remaining spokes. As a result, deviations from a hub-spoke network to global free trade cannot be farsightedly deterred, and global free trade via three country PTAs remains the unique stable outcome absent plurilateral agreements.

Figure 2 presents the LCS when plurilateral agreements are feasible.[21] Consistent with Section 4.1.2, the LCS includes both three country PTAs and hub-plurilateral networks across most of the parameter space. However, the set of stable networks varies with the elasticity of substitution σ, reflecting how the strength of MFN-based liberalization in the differentiated goods sector reshapes countries' incentives.

When σ is sufficiently smaller than the special case of $\sigma = 2$, the unique element of the LCS is global free trade via three-country PTAs. In this region, hub-plurilateral networks $g_{i,jk}^{H-PL}$ are not stable. Intuitively, when varieties are weak substitutes, the consumer gains from MFN tariff reductions in the differentiated goods sector are small. As a result, spokes do not benefit sufficiently from forming a plurilateral agreement, and equation (23) fails to hold. In this case, the hub country cannot credibly threaten a farsighted sequence of

[21] The figure assumes $N = 25$ firms. Appendix Figure A.1 reports analogous results for $N = 5$. Qualitatively, these two cases cover the broad LCS structure for much of the parameter space.

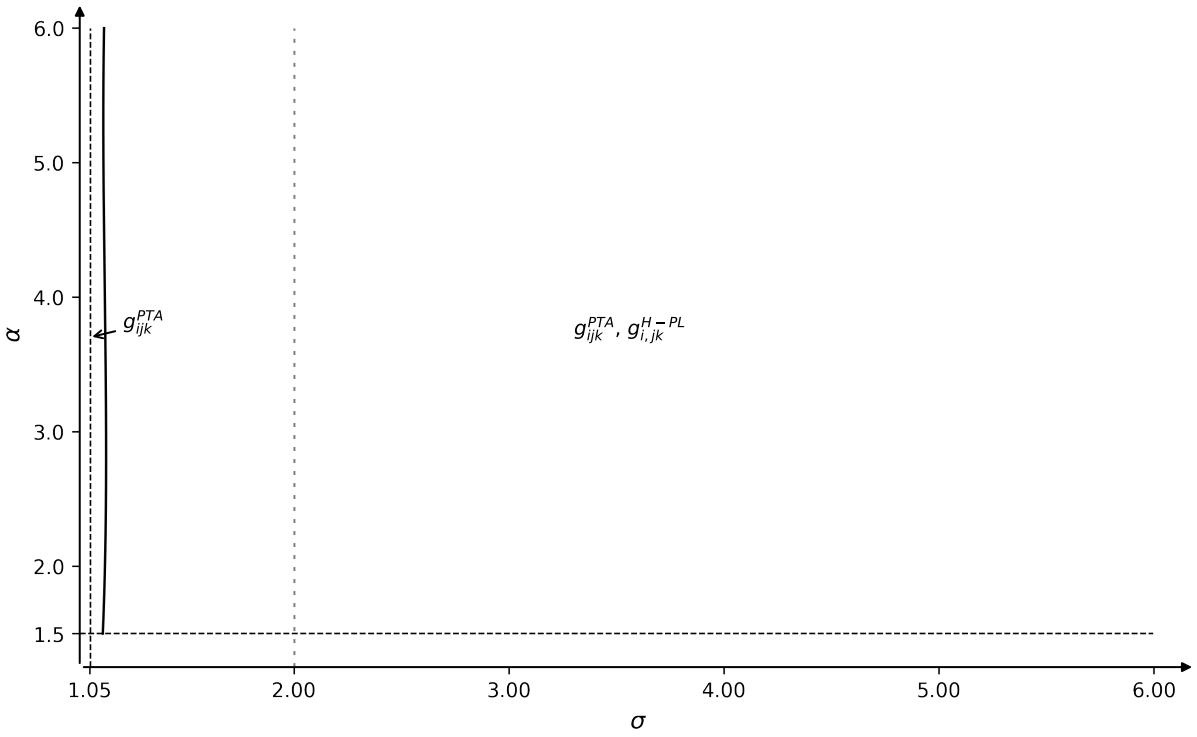


Figure 2: LCS with Symmetric Countries

Notes: Symmetric countries. $N = 25$ firms. Market size $\alpha \geq 1.5$ measured on the y-axis. Elasticity of substitution $\sigma \geq 1.05$ measured on the x-axis.

deviations that would deter the spokes from deviating from a hub–spoke network to a three-country FTA. Consequently, hub-plurilateral networks drop out of the LCS.

Overall, however, the insights from the special case $\sigma = 2$ extend to the general environment. Plurilateral agreements do not replace preferential liberalization, nor do they destabilize global free trade when it is attainable. Instead, they reshape the strategic environment by creating additional MFN-based margins of cooperation. When the gains from plurilateral liberalization are sufficiently large, these margins expand the set of stable trade agreement networks, allowing hybrid architectures that combine preferential access in traditional sectors with MFN liberalization in the differentiated goods sector.

5 Stable Trade Networks with Exclusion Incentives

We next consider environments with exclusion incentives, where asymmetries across countries prevent the expansion of PTAs to include all trading partners. In contrast to the symmetric benchmark, PTAs—as an institution on its own—are no longer sufficient to de-

liver global free trade in equilibrium. Instead, dominant PTA members have an incentive to preserve preferential rents by excluding weaker partners from further liberalization.

This section shows that plurilateral agreements play a distinct and economically meaningful role precisely in such constrained environments. While exclusion incentives continue to determine which countries exchange preferential access, plurilateral agreements expand the scope of liberalization by introducing MFN-based cooperation in selected sectors. As a result, plurilateral liberalization does not eliminate exclusion incentives or overturn the emergence of PTAs, but it enables additional global liberalization once the institutional limits of preferential liberalization bind.

We now introduce asymmetric supply-side characteristics in the homogeneous goods sector. Specifically, we assume that two countries, denoted s and s', have smaller endowments of their comparative advantage goods than the third country l, so that $e_s = e_{s'} < e_l$. All other parameters—including demand conditions and the structure of the differentiated goods sector—remain symmetric across countries.

This asymmetry has two related implications for PTA formation. First, smaller countries are more attractive PTA partners. Because all foreign countries account for the same share of world imports, they provide identical export market access. But, smaller countries account for a smaller share of world exports of non-numéraire goods and therefore pose a weaker import competition threat. As a result, exchanging preferential access with a small country yields relatively high net gains. Second, when there are two small countries and one large country, these incentives generate an exclusion motive: the two small countries prefer to form a bilateral PTA with each other rather than expand the agreement to include the large country, as doing so would dilute the preferential rents they enjoy in each other's markets. This environment therefore provides a natural setting in which preferential liberalization becomes constrained.

5.1 The Special Case: $\sigma = 2$

5.1.1 Constrained Institutional Framework: No Plurilaterals

In the absence of plurilateral agreements, exclusion incentives constrain further trade liberalization. The two small countries strictly prefer to form a bilateral CU and exclude the large country from further CU expansion. As a result, PTAs alone are no longer sufficient to deliver global free trade in this environment.

Expanding a bilateral PTA to include the large country would substantially increase import competition in the small countries' markets while delivering limited additional export gains, thereby diluting the preferential rents that the small countries enjoy in each

other's markets. Consequently, the small countries have no incentive to expand their CU, and the large country is unable to induce expansion unilaterally. Moreover, from any alternative trade network, the two small countries can profitably deviate to form their bilateral CU, and such deviations cannot be farsightedly deterred.

Proposition 3 summarizes this outcome.

Proposition 3 *Suppose $\sigma = 2$ and $N \geq 5$. Then, the small countries prefer their bilateral CU to all other trade agremeent networks when there are two small countries and one large country: $g^{CU}_{ss'} \succ_s g$ for all $g \neq g^{CU}_{ss'}$. Given these preferences, the LCS consists of the CU between the two small countries, $g^{CU}_{ss'}$.*

This outcome does not reflect a generic instability of global free trade. Rather, it arises from the institutional nature of PTAs: expanding a PTA necessarily reallocates discriminatory access and dilutes incumbent members' rents. Once this constraint binds, institutional rigidity—rather than a lack of gains from trade—becomes the binding obstacle to further liberalization.

5.1.2 Extended Institutional Framework: Plurilaterals

Allowing plurilateral agreements fundamentally changes this conclusion by separating the identity of PTA partners from the scope of tariff liberalization. PTAs continue to determine which countries exchange preferential access, but plurilateral agreements allow countries to expand liberalization on an MFN basis in the differentiated goods sector without altering existing PTAs. As a result, exclusion incentives remain binding in the preferential dimension, but no longer prevent mutually beneficial global liberalization in selected sectors.

These effects are reflected in the following payoff rankings:

$$g^{PTA-PL}_{ss',lss'} \succ_s \{g^{CU}_{ss'}, g^{FTA-PL}_{ss',ls}\} \succ_s g \text{ for } g \notin \{g^{PTA-PL}_{ss',lss'}, g^{CU}_{ss'}, g^{FTA-PL}_{ss',ls}\} \tag{24}$$

$$g \succ_l g^{PTA}_{ss'} \text{ for } g \in \{g^{PTA-PL}_{ss',lss'}, g^{FTA-PL}_{ss',ls}\} \tag{25}$$

$$g^{PTA-PL}_{ss',lss'} \succ_l g^{FTA-PL}_{ss',ls} \text{ if } g^{FTA-PL}_{ss',ls} \succ_s g^{CU}_{ss'}. \tag{26}$$

Equation (24) summarizes small country preferences: plurilaterals alter, but do not eliminate, the exclusion incentive. As PTA insiders, the small countries still prefer to exclude the large country from preferential access. However, they strictly prefer to augment their bilateral PTA with a three-country plurilateral agreement. The resulting PTA–plurilateral network is their most preferred outcome. Crucially, plurilateral agreements relax the binding constraint on further liberalization that existed in the absence of plurilaterals. The

small countries can preserve preferential rents via their PTA while simultaneously benefiting from MFN-based tariff reductions in the differentiated goods sector.

Equations (25)-(26) summarize the large country's preferences. Equation (25) shows that the large country strictly prefers participating in plurilateral agreements to remaining a PTA outsider, reflecting the gains from global MFN-based liberalization in the differentiated goods sector. Equation (26) governs the large country's incentives over alternative plurilateral-based networks. In particular, it describes situations where the large country does not benefit from destabilizing the small countries' most preferred outcome.

Taken together, these payoff rankings imply the following result.[22]

Proposition 4 *Suppose $\sigma = 2$ and $N \geq 5$. Then, the preference rankings in equations (24)-(26) are satisfied for the case of two small countries and one large country. Given these preferences, the LCS is the PTA–plurilateral network $g^{PTA-PL}_{ss',lss'}$, which consists of an FTA or CU between the two small countries and a three-country plurilateral agreement.*

Comparing Propositions 3 and 4 highlights the role of plurilateral agreements in environments where PTA formation incentives constrain the attainment of global free trade. When dominant countries wish to exclude others from preferential access, they are unwilling to eliminate tariffs across all sectors with non-members. Plurilateral agreements do not overturn this exclusion incentive. Instead, they provide an alternative institutional margin along which cooperation can expand, allowing MFN-based liberalization in selected sectors even when PTA expansion is exhausted.

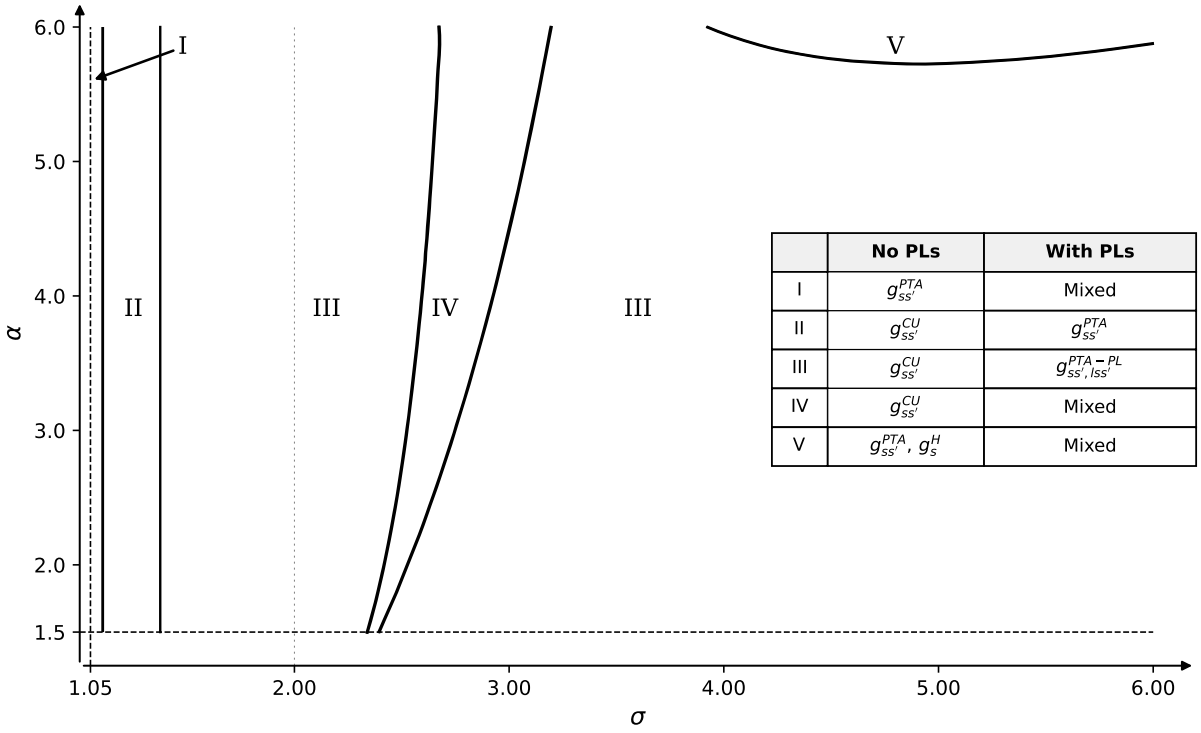
The stability of the PTA–plurilateral network follows from farsighted reasoning. The small countries have no incentive to deviate, as the network preserves preferential rents while delivering additional liberalization. Any unilateral deviation by the large country is deterred by a sequence of farsighted coalitional deviations that restores the PTA–plurilateral network. Finally, from any alternative trade network, there exists a profitable deviation—typically by the small countries to their bilateral PTA, or by a broader coalition toward the PTA–plurilateral network—that cannot be deterred. As a result, the PTA-plurilateral network $g^{PTA-PL}_{ss',lss'}$ constitutes the LCS.

5.2 Moving Beyond the Special Case

We now examine how the insights from the analytically tractable case extend to the broader parameter space. Figure 3 summarizes the LCS as a function of the elasticity of substitu-

[22] The CPNE mirrors the LCS in the special case under consideration.

tion, σ, and market size, α, for the case of two small countries and one large country when $N = 25$ firms.[23]



	No PLs	With PLs
I	$g_{ss'}^{PTA}$	Mixed
II	$g_{ss'}^{CU}$	$g_{ss'}^{PTA}$
III	$g_{ss'}^{CU}$	$g_{ss',ls's'}^{PTA-PL}$
IV	$g_{ss'}^{CU}$	Mixed
V	$g_{ss'}^{PTA}$, g_{s}^{H}	Mixed

Figure 3: LCS with Exclusion Incentives: The Impact of Plurilaterals

Notes: Two small countries and one large country. $N = 25$ firms. Summary table reports the LCS under the constrained (No PLs) and extended (With PLs) institutional frameworks. "Mixed" indicates the LCS contains both networks with and without plurilaterals. Market size $\alpha \geq 1.5$ measured on the y-axis. Elasticity of substitution $\sigma \geq 1.05$ measured on the x-axis.

When the institutional framework does not allow plurilateral agreements, Figure 3 shows that Proposition 3 extends to most of the parameter space. In regions II, III, and IV, the LCS consists solely of the bilateral CU between the small countries, $g_{ss'}^{CU}$, whereas in region I the PTA between the two small countries can also be an FTA. In all these regions, exclusion incentives are sufficiently strong that the small countries strictly prefer to preserve preferential rents by excluding the large country from further PTA expansion. As in the special case, global free trade cannot be sustained through PTAs alone.

This outcome changes only in a limited region of the parameter space. In region V, characterized by sufficiently large σ and α, a small country prefers to act as the hub in a hub-spoke network rather than remain a CU insider. In this region, preferential access

[23] Appendix Figure A.2 provides analogous results for $N = 5$ firms. Qualitatively, the two cases $N = 5$ and $N = 25$ cover the broad LCS structure for much of the parameter space.

to both spoke markets becomes sufficiently attractive to small countries that hub-spoke networks also enter the LCS. Even here, however, exclusion incentives persist and prevent the emergence of global free trade through PTAs alone.

Allowing plurilateral agreements substantially alters the set of stable trade networks. As shown in Figure 3, the result from Proposition 4 holds over a large portion of the parameter space (region III): the LCS consists of PTA–plurilateral networks that combine a bilateral PTA between the two small countries with a three-country plurilateral agreement. In these regions, the preference rankings derived in Section 5.1.2 continue to hold. Plurilateral agreements permit MFN-based liberalization in the differentiated goods sector while preserving the PTA network structure impied by exclusion incentives. Depending on the values of σ and α across this region, the shift to PTA-plurilateral networks closes 22% to 57% of the global welfare gap between $g_{ss'}^{CU}$ and the welfare-maximizing global free trade network, a significant movement towards global efficiency.

In other regions, the LCS differs because countries' preferences over trade networks change. In region II, the gains from plurilateral liberalization are too small to outweigh the exclusion incentive. As a result, the small countries' most preferred outcome remains their bilateral CU, and plurilateral agreements do not arise in equilibrium.

In region IV, when σ lies in an intermediate range, additional networks enter the LCS. In this region, FTA–plurilateral networks become sufficiently attractive that the large country is willing to trigger deviations that were previously deterred. As a result, the LCS expands to include a mix of networks involving combinations of PTAs and plurilateral agreements. A similar multiplicity arises at the boundaries of the parameter space: in regions I and V, PTA and PTA–plurilateral networks coexist in the LCS. Importantly, even in these cases, plurilateral agreements continue to coexist with PTAs rather than replace them.

Taken together, Figure 3 shows that the qualitative insights from the analytically tractable case of $\sigma = 2$ extend broadly across the parameter space. In the absence of plurilateral agreements, exclusion incentives prevent PTAs from delivering global free trade. Allowing plurilateral agreements does not remove this constraint or alter which countries trade preferentially. Instead, plurilateral agreements introduce an additional institutional margin along which cooperation can expand, enabling MFN-based liberalization in selected sectors even when PTA expansion is exhausted. In this sense, plurilateral agreements do not change which countries trade preferentially; they change how much liberalization is ultimately achieved. This robustness beyond the special case underscores that plurilateral agreements operate as a general mechanism for relaxing binding institutional constraints on trade liberalization, rather than as a knife-edge outcome tied to analytical tractability.

6 Stable Trade Networks with Free Riding Incentives

We now turn to environments characterized by free-riding incentives. In contrast to exclusion incentives—where PTA insiders resist expansion to preserve preferential rents—free-riding incentives arise when a country has little incentive to participate in preferential liberalization undertaken by other countries. In such settings, PTA expansion becomes constrained not because insiders block entry, but because the free-riding country refuses to participate. As in the case of exclusion incentives, the central question is whether plurilateral agreements can relax these institutional constraints by expanding MFN-based liberalization while leaving preferential access unchanged.

To generate free-riding incentives, we introduce an alternative asymmetry in the homogeneous goods sector. One country, denoted s, has a smaller endowment of its comparative advantage good than the remaining two countries l and l', so that $e_s < e_l = e_{l'}$. All other parameters—including demand conditions and the structure of the differentiated goods sector—remain symmetric across countries. Due to its limited supply capacity, the small country derives relatively modest export gains from preferential access to large markets while facing substantial import competition under reciprocal liberalization. At the same time, it benefits from tariff complementarity following PTA formation by the large countries, as PTA members endogenously reduce their external tariffs on non-members. As a result, the small country prefers to remain outside large-country PTAs, accepting some discrimination while free riding on partners' liberalization.

6.1 The Special Case: $\sigma = 2$

6.1.1 Constrained Institutional Framework: No Plurilaterals

We now analyze the special case of $\sigma = 2$ with two large countries, l and l', and one small country, s. In this environment, the free-riding incentive described above governs equilibrium PTA formation.

If the small country refuses to form a bilateral PTA with a large country, the large countries benefit from their own PTA. Moreover, they prefer CU rather than FTA formation as coordinating their policy vis-à-vis the outsider internalizes the externalities they otherwise impose on each other via tariff complementarity. As a result, the relevant comparison for the small country is between remaining a CU outsider and becoming the hub in a hub-spoke network, as the latter is typically the preferred network for an FTA member. While being the hub would grant the small country exclusive preferential access to both foreign markets, the free-riding incentive is sufficiently strong that the small country

strictly prefers remaining a CU outsider. The relevant payoff rankings therefore satisfy:

$$g_{ll}^{FTA} \succ_s \left\{g_{ll}^{CU}, g_{\emptyset}\right\} \succ_s g \quad \text{for all } g \notin \left\{g_{ll'}^{FTA}, g_{ll'}^{CU}, g_{\emptyset}\right\} \tag{27}$$

$$g_{ll'}^{CU} \succ_l g_{ll'}^{FTA} \succ_l g_{\emptyset}. \tag{28}$$

We can now summarize the LCS.

Proposition 5 *Suppose $\sigma = 2$ and $N \geq 25$. Then, the preferences in equations (27)–(28) hold for the case of two large countries and one small country. Given these preferences, the unique element of the LCS is the CU between the large countries, $g_{ll'}^{CU}$.*

The logic underlying Proposition 5 follows directly from the free-riding incentive. The small country will not participate in any deviation that leaves it as a PTA member, ruling out expansions of the large-country PTA. At the same time, the large countries prefer coordinating their external tariffs through a CU. Thus, any deviation away from $g_{ll'}^{CU}$ is either unprofitable or farsightedly deterred by a return to the CU.

No other trade network can be stable. Because the small country benefits from deviating by exiting any PTA it has formed regardless of the PTA the large countries may or may not form, such deviations cannot be farsightedly deterred. In combination with the large country's preference for a CU, this also implies the large country deviation to form a CU cannot be farisghtedly deterred from either their FTA or a situation of no agreements. Thus, in the absence of plurilateral agreements, free-riding incentives constrain further liberalization beyond $g_{ll'}^{CU}$.

6.1.2 Extended Institutional Framework: Plurilaterals

Allowing plurilateral agreements reshapes the set of stable trade networks by creating an additional margin of cooperation that does not require reciprocal preferential liberalization. As before, PTAs continue to determine who trades preferentially, while plurilateral agreements expand liberalization along MFN-based sectoral margins. In particular, plurilaterals permit MFN-based tariff reductions in the differentiated goods sector while leaving the PTA structure intact—a feature that is especially consequential in environments with free-riding incentives.

The relevant welfare rankings over trade networks are

$$g_{ll',\cdot}^{FTA\text{-}PL} \succ_s g_{ll'}^{FTA} \succ_s g_{ll',sll'}^{CU\text{-}PL} \succ_s \left\{g_{ll'}^{CU}, g_{\emptyset}, g_{\cdot}^{PL}\right\} \tag{29}$$

$$g_{ll',sll'}^{CU\text{-}PL} \succ_l \left\{g_{ll',\cdot}^{FTA\text{-}PL}, g_{ll'}^{PTA}\right\} \succ_l g_{\cdot}^{PL} \succ_l g_{\emptyset}, \tag{30}$$

where the subscript dot in $g_{\cdot}^{PL}$ and $g_{ll',\cdot}^{FTA\text{-}PL}$ indicates that any two countries or all three countries could form a plurilateral. Despite its free riding incentive, the small country benefits from forming a plurilateral agreement following PTA formation by the large countries. Doing so allows the small country to benefit from reciprocal tariff reductions in the differentiated goods sector while avoiding the costly reciprocal liberalization that deters its PTA participation. Since all countries are symmetric in the differentiated good sector, the large countries also benefit from plurilateral liberalization, particularly when combined with their bilateral CU.

We can now summarize the LCS in this case.[24]

Proposition 6 *Suppose $\sigma = 2$ and $N \geq 25$. Then, the preferences in equations (27)–(30) hold for the case of two large countries and one small country. Given these preferences, the unique element of the LCS is the CU–plurilateral network $g_{ll',sll'}^{CU\text{-}PL}$.*

The logic parallels Proposition 5, with one crucial difference. As before, free-riding incentives continue to block PTA participation by the small country. But, plurilateral agreements now allow MFN-based liberalization to proceed without altering preferential access. As a result, the CU between the large countries is augmented—but not replaced—by a three-country plurilateral agreement. Any deviation away from $g_{ll',sll'}^{CU\text{-}PL}$ is either unprofitable or farsightedly deterred by a sequence of deviations that restores the CU–plurilateral network. And, by similar logic, no alternative trade network can be stable because there is a deviation that cannot be deterred: either the small country exits its PTAs or the large countries form their CU.

Taken together, Proposition 6 shows that plurilateral agreements relax the constraints imposed by free-riding incentives without eliminating them. PTAs continue to determine who trades preferentially, while plurilaterals determine how much liberalization is ultimately achieved.

6.2 Moving Beyond the Special Case

We now examine how the results from the analytically tractable case extend across the parameter space. Figure 4 illustrates how the LCS varies with the elasticity of substitution σ when market size is fixed at $\alpha = 5.05$ in the institutional environment without plurilateral agreements. While fixing $\alpha = 5.05$ is useful for transparently illustrating these reversals, Appendix Figures A.3-A.4 show that the qualitative structure of the LCS extends more

[24]The CPNE mirrors the LCS in this special case under consideration.

generally across the parameter space. Payoff rankings at the top of the figure indicate the preference reversals that occur at the threshold values of σ where the LCS changes.

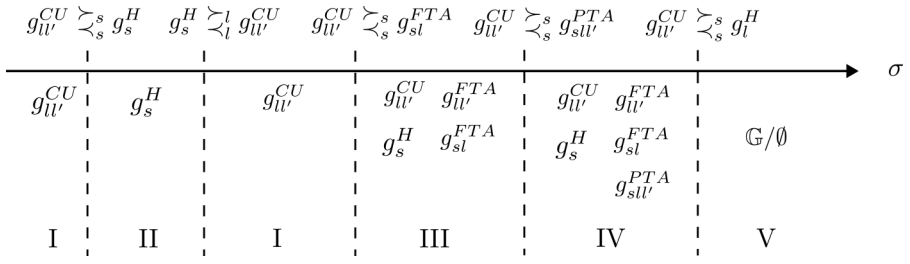


Figure 4: LCS with Free Riding Incentive and Varying σ: Without Plurilaterals

Notes: Two large countries and one small country. $N = 25$ firms. Market size fixed at $\alpha = 5.05$. σ measured on the line. Payoff rankings at top indicate the preference reversal at the threshold σ: $g \succsim_i^{\prec} g'$ denotes $g \succ_i g'$ ($g' \succ_i g$) below (above) the threshold σ. $\sigma = 2$ lies in region I where $g_{ll'}^{CU} \succ_s g_s^H$.

In the absence of plurilateral agreements, the LCS reflects the strength of the free-riding incentive as σ varies. For rather low and then again for moderate values of σ (i.e. region I), the small country prefers to remain outside PTAs, and the LCS consists of a CU between the two large countries, consistent with Proposition 5. As σ increases, the differentiated goods sector becomes more important, weakening the free-riding incentive. Once σ is sufficiently large, the small country prefers bilateral preferential access, violating equation (27), and the LCS expands to include bilateral FTAs. For even larger values of σ, additional networks enter the LCS as further preference rankings reverse.

Allowing plurilateral agreements fundamentally alters how free-riding incentives shape equilibrium trade networks. While free riding continues to discourage participation in PTAs, plurilaterals enable MFN-based liberalization in the differentiated goods sector without requiring reciprocal preferential liberalization in all sectors. As a result, the free-riding constraint no longer blocks global tariff liberalization, even when PTA participation remains unattractive.

Figure 5 shows how the LCS varies with σ when plurilateral agreements are permitted (we again consider $\alpha = 5.05$ for illustration). In region II, where σ is sufficiently low, the gains from plurilateral liberalization are limited and the large countries prefer to remain CU insiders rather than form a three-country plurilateral agreement. In this region, the LCS coincides with the PTA-only case. As σ increases into region III, plurilateral liberalization becomes attractive and the CU–plurilateral network emerges as the unique element of the LCS, mirroring the $\sigma = 2$ benchmark in Proposition 6. For intermediate values of σ, the LCS in regions I include both networks with and without plurilaterals, reflecting competing incentives between preferential access and MFN liberalization. As σ increases

further into regions IV and I, additional networks—such as hub-plurilateral and FTA–plurilateral configurations—enter the LCS as the gains from MFN liberalization in the differentiated sector dominate the remaining free-riding incentives.

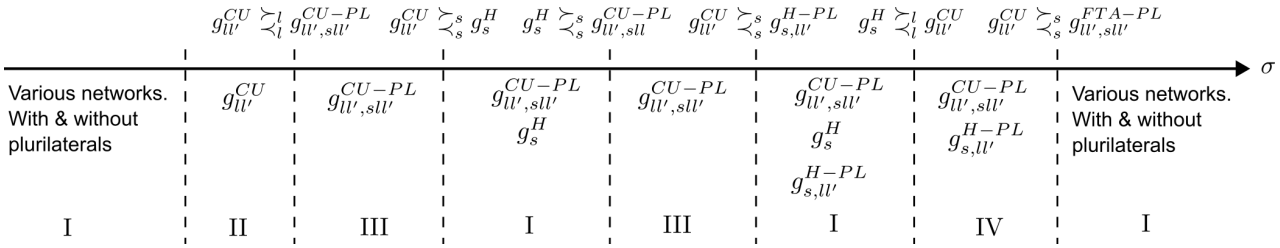


Figure 5: LCS with Free Riding Incentive and Varying σ: With Plurilaterals

Notes: Two large countries and one small country. $N = 25$ firms. Market size fixed at $\alpha = 5.05$. σ measured on the line. Payoff rankings at top indicate the preference reversal at the threshold σ: $g \succsim_i^{} g'$ denotes $g \succ_i g'$ ($g' \succ_i g$) below (above) the threshold σ. $\sigma = 2$ lies in region III where $g_{ll'}^{CU} \succ_s g_s^H$. Multiple regions labeled I to coincide with labels in Figure 6.

Figure 6 generalizes these results by varying α in addition to σ.[25] When σ is sufficiently low, plurilateral agreements do not arise in equilibrium and the LCS consists solely of a CU between the large countries. Above this lower bound, plurilateral agreements emerge as part of the LCS and coexist with PTAs across wide regions of the parameter space. Depending on the values of σ and α in these regions, the shift to PTA-plurilateral networks bridges 31% to 67% of the global welfare gap between $g_{ll'}^{CU}$ and the welfare-maximizing global free trade network, a significant movement towards global efficiency. Thus, plurilateral agreements do not change which countries trade preferentially; they change how much liberalization is ultimately achieved. This robustness beyond the special case underscores that plurilateral agreements operate as a general mechanism for relaxing binding institutional constraints on trade liberalization and improving global welfare.

Ultimately, the insights from Propositions 5 and 6 extend well beyond the analytically tractable case of $\sigma = 2$. While the composition of the LCS varies with σ and α, the underlying mechanism is robust: free-riding incentives can block participation in PTAs, but plurilateral agreements open an additional institutional margin that enables MFN-based liberalization in the differentiated goods sector. In this sense, $\sigma = 2$ provides a transparent benchmark that captures a general mechanism rather than a knife-edge case. Across a wide range of parameter values, plurilateral agreements mitigate free-riding constraints and expand the scope for global tariff liberalization without displacing existing PTAs.

[25] See Appendix Figure A.5 for analogous results when $N = 5$ firms. Qualitatively, the two cases $N = 5$ and $N = 25$ cover the broad LCS structure for much of the parameter space.

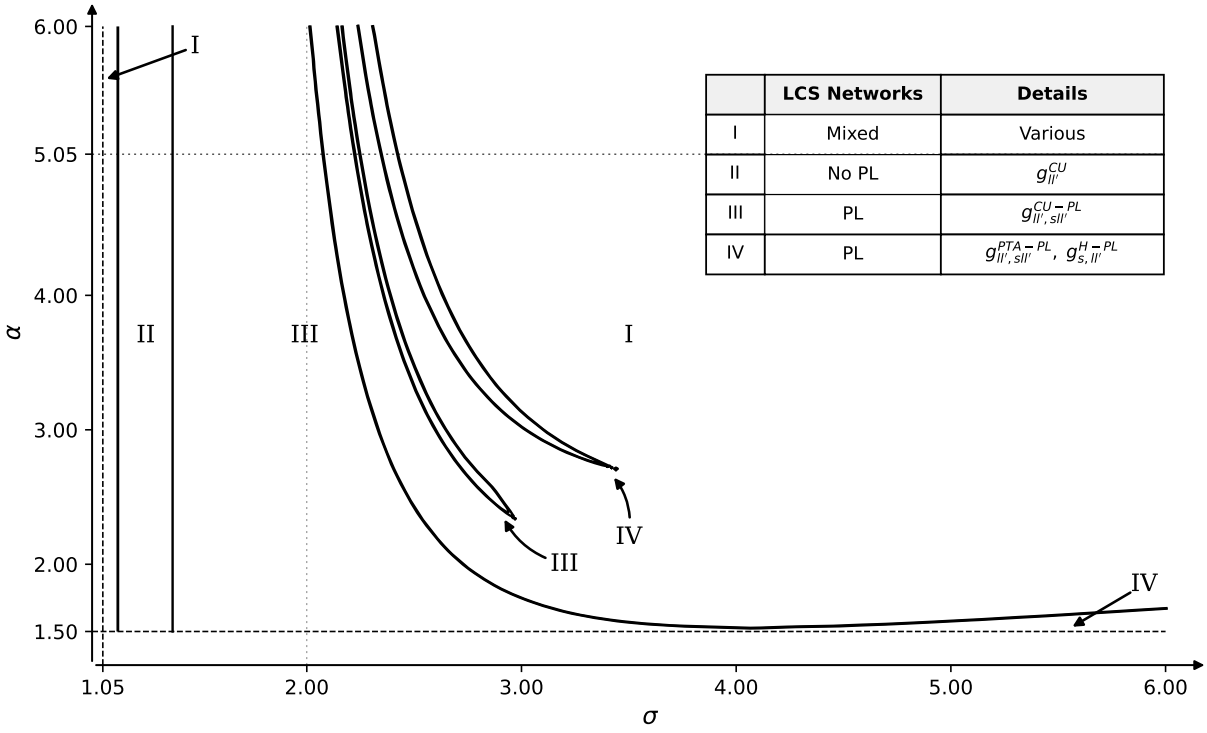


Figure 6: LCS with Free Riding Incentive and Varying (α, σ): With Plurilaterals

Notes: Two large countries and one small country. $N = 25$ firms. Summary table reports LCS under the extended institutional framework (With PLs). "Mixed" indicates LCS contains networks with and without plurilaterals. Market size $\alpha \geq 1.5$ measured on the y-axis. Elasticity of substitution $\sigma \geq 1.05$ measured on the x-axis. Appendix Figure A.3 shows LCS for the constrained framework without plurilaterals.

7 Conclusion

We study the role of plurilateral agreements in an environment where multilateral liberalization has stalled and PTAs have become the dominant vehicle for tariff liberalization. Using a model of endogenous trade agreement formation with farsighted governments, we show that allowing liberalization on an MFN basis within specific sectors alters the set of stable trade agreement networks in a systematic way: plurilateral agreements expand the scope of liberalization without undermining the PTAs that would otherwise arise.

A central finding of the analysis is that plurilateral agreements do not replace or destabilize existing PTAs. Across the environments we consider, PTAs remain part of the equilibrium trade architecture. Instead, plurilateral agreements emerge as complements that augment the stable network of trade agreements by adding MFN-based sectoral liberalization while leaving preferential access unchanged. This coexistence result is non-trivial and helps allay concerns—frequently raised in policy discussions—that plurilateral initiatives

may weaken or fragment the existing trade system.

A key contribution of the paper is clarifying why plurilateral agreements matter precisely when preferential liberalization reaches its limits. PTAs are shown to be an effective but rigid institutional response to multilateral deadlock: by bundling liberalization across sectors and relying on discrimination, they facilitate reciprocity and limit fragmentation. But, these features imply that once preferential rents are fully exploited, further liberalization through PTAs is constrained by exclusion or free-riding incentives. In such environments, plurilateral agreements relax binding constraints by allowing countries to liberalize in sectors of mutual interest without requiring further expansion of preferential access.

In settings where global free trade is unattainable through PTAs alone, plurilateral agreements facilitate additional global tariff reductions by operating as an institutional overlay on existing PTA networks. The equilibrium outcome with plurilaterals corresponds to the PTA-based outcome that would arise absent plurilaterals, augmented by MFN-based sectoral liberalization. Plurilaterals thus expand the frontier of feasible cooperation rather than substituting for existing institutions.

Taken together, the analysis offers a coherent interpretation of recent developments in the global trading system. The evolution from multilateralism to PTAs and, increasingly, to plurilateral initiatives need not be viewed as fragmentation. Instead, it reflects a process of sequential institutional adaptation in which new forms of cooperation arise once existing institutional possibilities reach their limits. PTAs represent a natural starting point in a post-multilateral world, but their rigidity eventually constrains further liberalization. Plurilateral agreements emerge at this stage as a flexible mechanism that sustains additional cooperation while preserving the preferential core of the trading system.

Several extensions offer promising directions for future research, including richer sectoral heterogeneity, the interaction between plurilateral agreements and non-tariff measures, and dynamic considerations such as entry, enforcement, and uncertainty. In addition, the model developed here could be used to investigate other PTA-related questions, for instance the role of market structure in PTA formation. More broadly, the results suggest that the growing prominence of plurilateral agreements is neither anomalous nor destabilizing. Rather, it reflects the institutional limits of preferential liberalization and the need for flexibility once those limits are reached—a dynamic likely to shape the future of global trade cooperation.

References

Abbott, K. W. and D. Snidal (2000): "Hard and soft law in international governance," *International Organization*, 54, 421–456.

Bagwell, K. and R. Staiger (1999): "Regionalism and Multilateral Tariff Cooperation," in *International trade policy and the Pacific rim*, ed. by J. Pigott and A. Woodland, Macmillan.

Bagwell, K. and R. W. Staiger (2002): *The economics of the world trading system*, MIT press.

Baldwin, R. (1995): "A domino theory of regionalism," in *Expanding Membership of the EU*, ed. by R. Baldwin, P. Haaparanta, and J. Kiander, Cambridge University Press, 25–48.

Berens, S., L. Chochua, and G. Willmann (2021): "The Farsighted Stability of Global Trade Policy Arrangements," CESifo Working Paper 9213, CESifo, Center for Economic Studies and ifo Institute, Munich.

Bown, C. P. and M. A. Crowley (2016): "The empirical landscape of trade policy," *Handbook of Commercial Policy*, 1, 3–108.

Chochua, L. and I. Iodice (2025): "Tariff Cuts, Policy Uncertainty, and the Force of Many: The Impact of Plurilateral Agreements," *CESifo Working Paper No. 12125*.

Chwe, M. S.-Y. (1994): "Farsighted Coalitional Stability," *Journal of Economic Theory*, 63, 299 – 325.

Davis, C. L. (2004): "International institutions and issue linkage: Building support for agricultural trade liberalization," *American Political Science Review*, 98, 153–169.

Facchini, G., P. Silva, and G. Willmann (2012): "The Customs Union issue: Why do we observe so few of them?" *Journal of International Economics*, 90, 136–147.

——— (2021): "The political economy of preferential trade agreements: An empirical investigation," *The Economic Journal*, 131, 3207–3240.

Freund, C. and E. Ornelas (2010): "Regional trade agreements," *Annual Review of Economics*, 2, 139–166.

Gnutzmann-Mkrtchyan, A. and C. Henn (2018): "Peeling away the layers: Impacts of durable tariff elimination," *Journal of International Economics*, 115, 259–276.

Hoekman, B. and C. Sabel (2019): "Open plurilateral agreements, international regulatory cooperation and the WTO," *Global Policy*, 10, 297–312.

——— (2021): "Plurilateral cooperation as an alternative to trade agreements: Innovating one domain at a time," *Global Policy*, 12, 49–60.

Hoekman, B. M. and P. C. Mavroidis (2015a): "Embracing diversity: Plurilateral agreements and the trading system," *World Trade Review*, 14, 101–116.

——— (2015b): "WTO 'à la carte'or 'menu du jour'? Assessing the Case for More Plurilateral Agreements," *European Journal of International Law*, 26, 319–343.

——— (2017): "MFN Clubs and Scheduling Additional Commitments in the GATT: Learning from the GATS," *European Journal of International Law*, 28, 387–407.

Keohane, R. O. and D. G. Victor (2011): "The regime complex for climate change," *Perspectives on Politics*, 9, 7–23.

Krishna, P. (1998): "Regionalism and Multilateralism: A Political Economy Approach," *Quarterly Journal of Economics*, 113, 227–251.

Krishna, P. and J. Lake (2019): "Preferential Trade Agreements: Recent theoretical and empirical developments," *Oxford Research Encylopedia of Economics and Finance*.

Krugman, P. (1991): "The move toward free trade zones. Policy implications of trade and currency zones: A symposium sponsored by the Federal Reserve Bank of Kansas City," *Federal Reserve Bank of Kansas City, Kansas City*, 7–41.

Lake, J. (2017): "Free trade agreements as dynamic farsighted networks," *Economic Inquiry*, 55, 31–50.

——— (2019): "Dynamic formation of Preferential Trade Agreements: The role of flexibility," *Canadian Journal of Economics*, 52, 132–177.

——— (2026): "The evolution of trade barriers in the 21st century," in *Elgar Encyclopedia of International Trade*, ed. by D. Mitra and J. Priyaranjan, Edward Elgar.

Lake, J., M. Nken, and H. M. Yildiz (2020): "Tariff bindings and the dynamic formation of preferential trade agreements," *Journal of International Economics*, 122, 103279.

Lake, J. and S. Roy (2017): "Are global trade negotiations behind a fragmented world of "gated globalization"?" *Journal of International Economics*, 108, 117–136.

Lake, J. and H. M. Yildiz (2016): "On the different geographic characteristics of Free Trade Agreements and Customs Unions," *Journal of International Economics*, 103, 213–233.

Lewis, M. K. (2025): "Plurilateralism and Regional Trade Agreements," *Journal of World Investment & Trade*, 25.

Maggi, G. (2014): "International trade agreements," in *Handbook of International Economics*, ed. by G. Gopinath, E. Helpman, and K. Rogoff, Elsevier, vol. 4, 317–390.

——— (2016): "Issue linkage," in *Handbook of Commercial Policy*, Elsevier, vol. 1, 513–564.

Mann, C. and X. Liu (2009): "The information technology agreement: Sui generis or model stepping stone," *Multilateralizing regionalism: Challenges for the global trading system*, 182–216.

Missios, P., K. Saggi, and H. M. Yildiz (2016): "External trade diversion, exclusion incentives and the nature of preferential trade agreements," *Journal of International Economics*, 99, 105–119.

Ornelas, E. (2005a): "Endogenous free trade agreements and the multilateral trading system," *Journal of International Economics*, 67, 471–497.

——— (2005b): "Rent destruction and the political viability of free trade agreements," *The Quarterly Journal of Economics*, 1475–1506.

——— (2005c): "Trade Creating Free Trade Areas and the Undermining of Multilateralism," *European Economic Review*, 49, 1717–1735.

——— (2008): "Feasible multilateralism and the effects of regionalism," *Journal of International Economics*, 74, 202–224.

Saggi, K., A. Woodland, and H. M. Yildiz (2013): "On the relationship between preferential and multilateral trade liberalization: The case of customs unions," *American Economic Journal: Microeconomics*, 5, 63–99.

Saggi, K. and H. M. Yildiz (2010): "Bilateralism, multilateralism, and the quest for global free trade," *Journal of International Economics*, 81, 26–37.

Tang, X. and R. Lascari (2017): *20 Years of the Information Technology Agreement: Boosting Trade, Innovation and Digital Connectivity*, World Trade Organization.

Winslett, G. (2018): "Critical Mass Agreements: The Proven Template for Trade Liberalization in the WTO," *World Trade Review*, 17, 405–426.

Wolfe, R. (2015): "First diagnose, then treat: What ails the Doha Round?" *World Trade Review*, 14, 7–28.

Wolff, A. W. (2023): *Revitalizing the world trading system*, Cambridge University Press.

A Additional Tables and Figures

Network	Countries	Optimal Formula
$g_{\emptyset}$	All (i)	$t_{ij}(g_{\emptyset}) = \frac{1}{8}(e_j + e_k)$
g_{ij}^{FTA}	Insiders (i,j)	$t_{ik}(g_{ij}^{FTA}) = \frac{1}{11}(5e_k - 4e_j)$
	Outsider (k)	$t_{ki}(g_{\emptyset}) = \frac{1}{8}(e_i + e_j)$
g_{ij}^{CU}	Insiders (i,j)	$t_{ik}(g_{ij}^{CU}) = \frac{1}{5}(2e_k - e_j)$
	Outsider (k)	$t_{ki}(g_{\emptyset}) = \frac{1}{8}(e_i + e_j)$
g_i^H	Spokes (j,k)	$t_{jk}(g_{ij}^{FTA}) = \frac{1}{11}(5e_k - 4e_i)$

Table A.1: Network-dependent Optimal External Tariffs for Homogeneous Goods Sector

Notes: Formulas derived from welfare maximization in competing exporters model. e_i denotes country i's endowment. Endowments constrained by $e_s/e_l \leq \frac{5}{3}$ to ensure non-negative exports across all trade agreement networks. All tariffs constrained by $0 \leq t_{ij}(g) \leq t_{ij}(g_{\emptyset})$. Lower bound excludes import subsidies. Upper bound imposed by WTO Article XXIV constraint that PTAs cannot raise external tariffs above pre-existing MFN levels. CU members jointly choose common external tariff $t_{ik}(g_{ij}^{CU}) = t_{jk}(g_{ij}^{CU})$. FTA members independently optimize external tariffs.

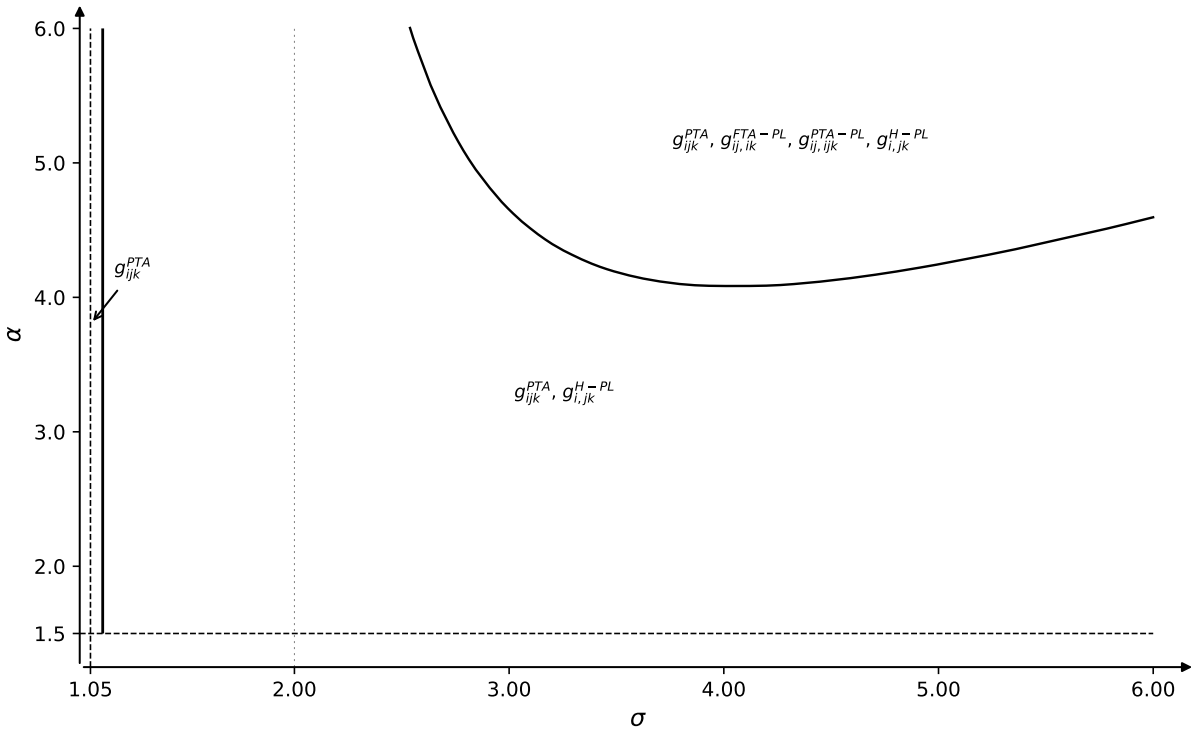


Figure A.1: LCS with Symmetric Countries

Notes: Symmetric countries. $N = 5$ firms. Market size $\alpha \geq 1.5$ measured on the y-axis. Elasticity of substitution $\sigma \geq 1.05$ measured on the x-axis. Without plurilaterals, g_{ijk}^{PTA} is the unique stable network for all parameter values.

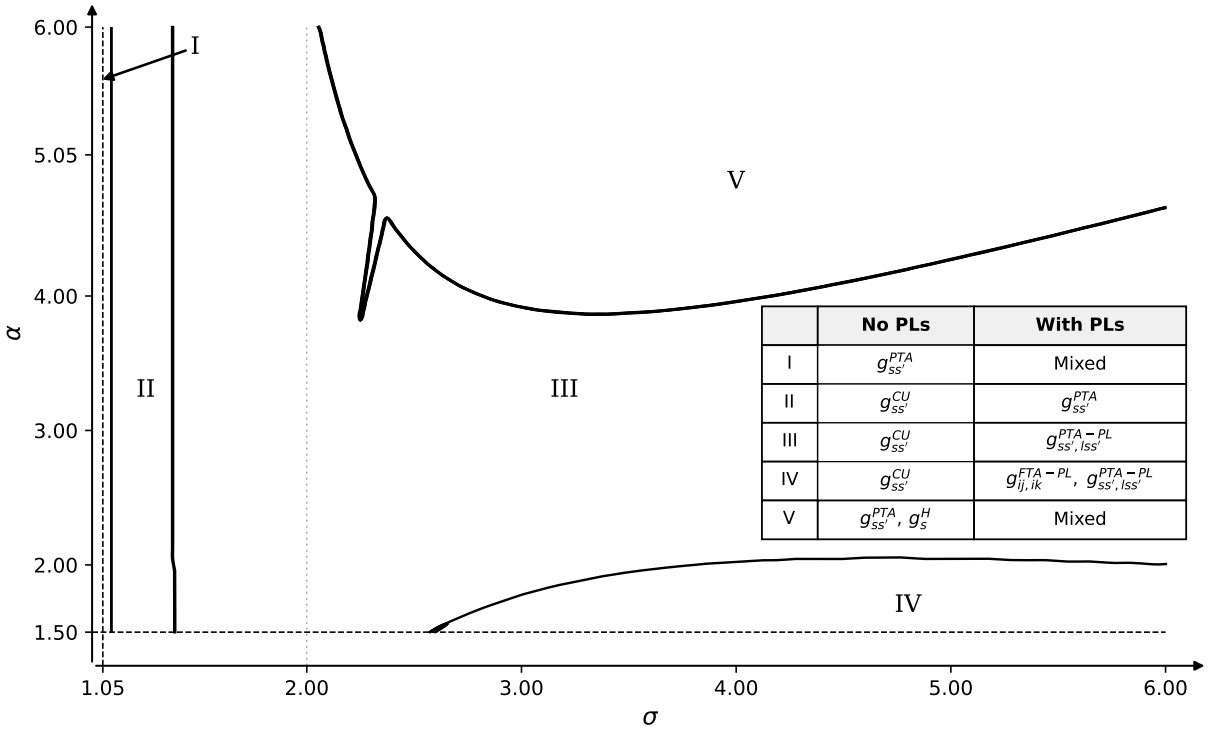


Figure A.2: LCS with Exclusion Incentives: The Impact of Plurilaterals

Notes: Two small countries and one large country. $N = 5$ firms. Summary table reports the LCS under the constrained (No PLs) and extended (With PLs) institutional frameworks. "Mixed" indicates the LCS contains both networks with and without plurilaterals. Market size $\alpha \geq 1.5$ measured on the y-axis. Elasticity of substitution $\sigma \geq 1.05$ measured on the x-axis.

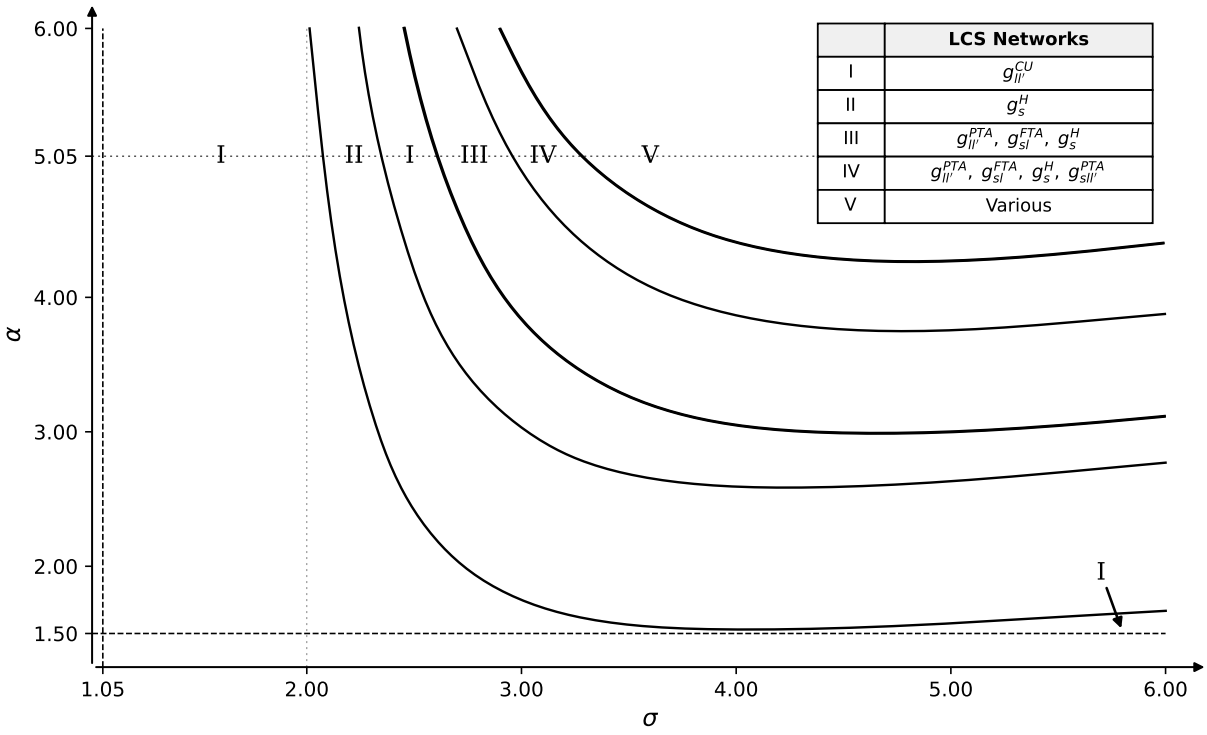


Figure A.3: LCS with Free Riding Incentive and Varying (α, σ): Without Plurilaterals

Notes: Two large countries and one small country. $N = 25$ firms. Market size $\alpha \geq 1.5$ measured on y-axis. Elasticity of substitution $\sigma \geq 1.05$ measured on x-axis. For more details about the LCS when $\alpha = 5.05$, see Figure 4 in main text.

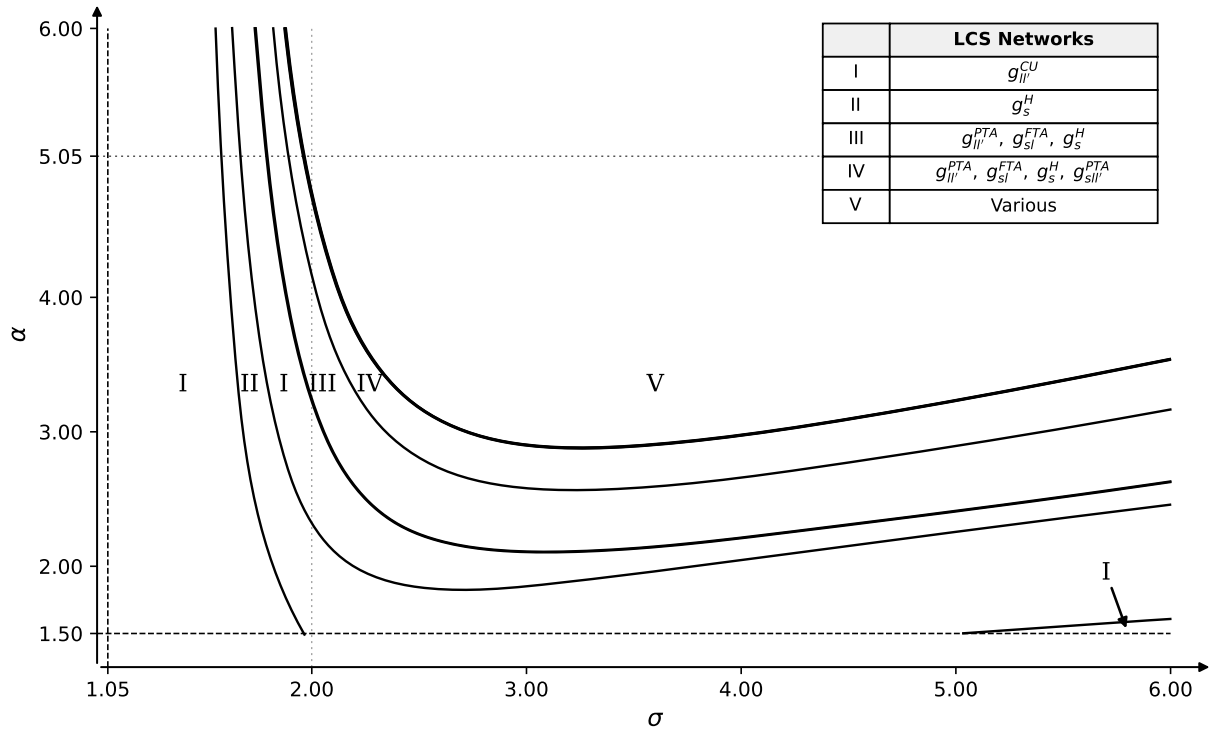


Figure A.4: LCS with Free Riding Incentive and Varying (α, σ): Without Plurilaterals

Notes: Two large countries and one small country. $N = 5$ firms. Market size $\alpha \geq 1.5$ measured on the y-axis. Elasticity of substitution $\sigma \geq 1.05$ measured on the x-axis.

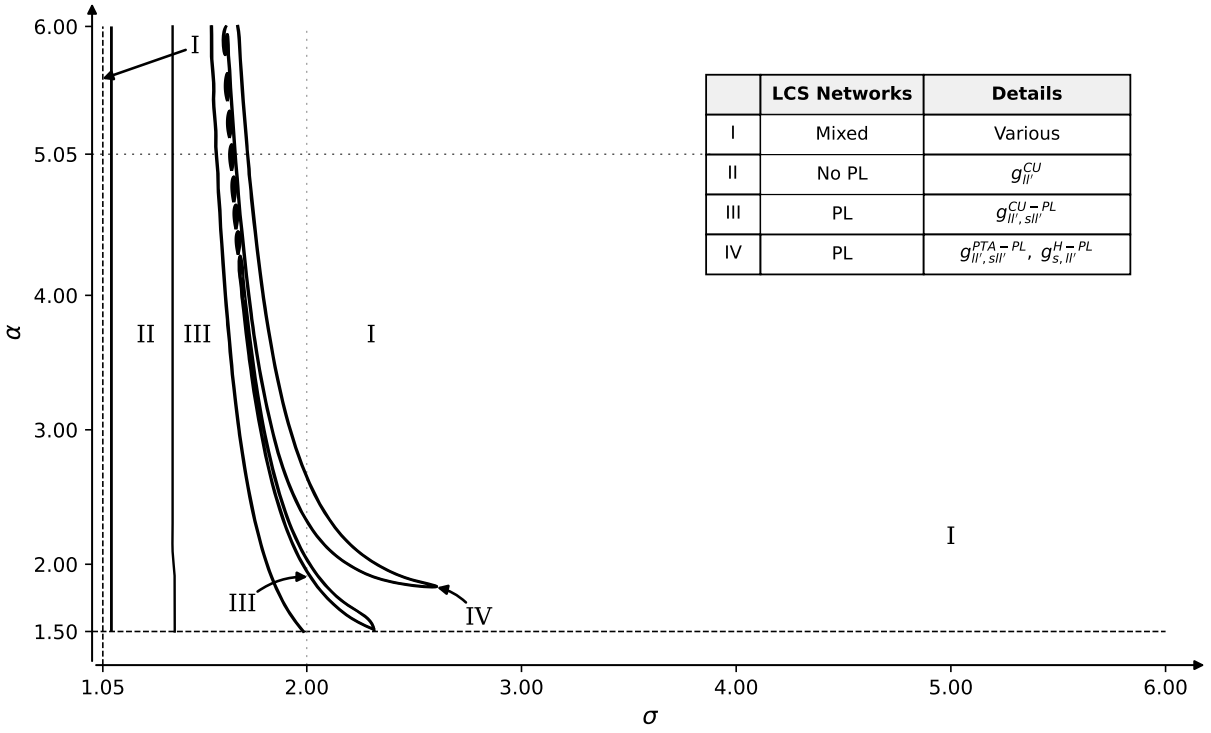


Figure A.5: LCS with Free Riding Incentive and Varying (α, σ): With Plurilaterals

Notes: Two large countries and one small country. $N = 5$ firms. Summary table reports LCS under the extended institutional framework (With PLs). "Mixed" indicates LCS contains both networks with and without plurilaterals. Market size $\alpha \geq 1.5$ measured on the y-axis. Elasticity of substitution $\sigma \geq 1.05$ measured on the x-axis.

B Welfare Expressions

Homogeneous Goods Sector. The following equations provide expressions for consumer surplus (CS), producer surplus (PS) and tariff revenue (TR) for country i in the homogeneous goods sector as a function of market size (α), endowments (e_i, e_j, e_k) and all bilateral tariffs ($t_{zz'}$) imposed by country z on country z'.

$$CS_i = \frac{1}{18}\left[(e_j + e_k - t_{ij} - t_{ik})^2 + (e_i + e_k + 2t_{ji} - t_{jk})^2 + (e_i + e_j + 2t_{ki} - t_{kj})^2\right] \tag{31}$$

$$PS_i = \frac{1}{3}e_i\left(6\alpha - 2e_i - e_j - e_k - 2t_{ji} + t_{jk} - 2t_{ki} + t_{kj}\right) \tag{32}$$

$$TR_i = \frac{1}{3}\left[e_k(2t_{ik} - t_{ij}) + e_j(2t_{ij} - t_{ik}) - 2(t_{ij}^2 - t_{ij}t_{ik} + t_{ik}^2)\right] \tag{33}$$

Differentiated Goods Sector. The following equations provide expressions for consumer surplus (CS), producer surplus (PS) and tariff revenue (TR) for country i in the differentiated goods sector as a function of market size (α), elasticity of substitution (σ), number of firms in each country N, and all bilateral tariffs ($\tau_{zz'}$) imposed by country z on country z'.

$$CS_i = \frac{1}{2}\left(\alpha - \frac{\sigma}{\sigma - 1}\tau_i\right)^2 \tag{34}$$

$$PS_i = N\frac{1}{\sigma - 1}\left[\left(\alpha - \frac{\sigma}{\sigma - 1}\tau_i\right)\tau_i^\sigma + \left(\alpha - \frac{\sigma}{\sigma - 1}\tau_j\right)\frac{\tau_j^\sigma}{\tau_{ji}^\sigma} + \left(\alpha - \frac{\sigma}{\sigma - 1}\tau_k\right)\frac{\tau_k^\sigma}{\tau_{ki}^\sigma}\right] \tag{35}$$

$$TR_i = N\frac{\sigma}{\sigma - 1}\left(\alpha - \frac{\sigma}{\sigma - 1}\tau_i\right)\tau_i^\sigma\left(\frac{\tau_{ij} - 1}{\tau_{ij}^\sigma} + \frac{\tau_{ik} - 1}{\tau_{ik}^\sigma}\right) \tag{36}$$

where $\tau_z = \left[N\left(\tau_{zi}^{1-\sigma} + \tau_{zj}^{1-\sigma} + \tau_{zk}^{1-\sigma}\right)\right]^{\frac{1}{1-\sigma}}$ for $z \in \{i, j, k\}$.

C Proofs

Proof of Proposition 1.

The proof is given in the main text following Example 1. ■

Proof of Proposition 2.

Given the preference rankings in Proposition 2, the logic used in the proof of Proposition 1 establishes that (i) $g \notin G_1$ for $g \notin \{g^{CU}_{ijk}, g^{FTA}_{ijk}, g^{H-PL}_{i,jk}\}$ and (ii) $g \in G_1$ for $g \in \{g^{CU}_{ijk}, g^{FTA}_{ijk}\}$. What remains to show is that any deviation from $g^{H-PL}_{i,jk}$ is farsightedly deterred and hence $g^{H-PL}_{i,jk} \in G_1$. Because i cannot become better off from deviating, we now focus on deviations by $S \subseteq \{j,k\}$.

First, consider the deviation by the spokes $S = \{j,k\}$ from $g^{H-PL}_{i,jk}$ to g^{FTA}_{ijk}. This is farsightedly deterred by the sequence of deviations $g^{FTA}_{ijk} \to_i g^{H}_{j} \to_{ik} g_{\emptyset} \to_{ijk} g^{H-PL}_{i,jk}$ because (i) each deviating coalition along the sequence strictly prefers the eventual outcome $g^{H-PL}_{i,jk}$ to the prevailing network and (ii) the initially deviating coalition S is no better off because the sequence of deviations returns the network $g^{H-PL}_{i,jk}$ before the sequence of deviations.

Second, consider any other deviation by a coalition $S \subseteq \{j,k\}$ from $g^{H-PL}_{i,jk}$ to some g. For any such g, such as $g^{FTA-PL}_{ij,jk}$ or g^{CU}_{jk}, a country z can initiate the sequence of deviations $g \to_z g_{\emptyset} \to_{z,z',z''} g^{H-PL}_{z,z'z''}$ that leaves either j or k no better off than under $g^{H-PL}_{i,jk}$. This deters any such deviation by $S \subseteq \{j,k\}$ from $g^{H-PL}_{i,jk}$.

We have now established $G_1 = f(\mathbb{G}) = \{g^{CU}_{ijk}, g^{FTA}_{ijk}, g^{H-PL}_{i,jk}\}$. Indeed, $G_1 = f(G_1)$ and hence G_1 is the LCS. We have already explained above that any deviation by $S \subseteq \{j,k\}$ from $g^{H-PL}_{i,jk}$ is farsightedly deterred by some $g \in G_1$. And, given there is no network g' that any coalition S can deviate to from $g \in \{g^{CU}_{ijk}, g^{FTA}_{ijk}\}$ such that S is better off, part (i) of Definition 3 implies any deviation by S from g to g' is farsightedly deterred. ■

Proof of Proposition 3.

Note that s and s' prefer g^{CU}_{ss} over any other g and l cannot deviate from g^{CU}_{ss} to any other g unilaterally. This implies (i) the deviation by $S = \{s, s'\}$ from any $g \neq g^{CU}_{ss'}$ to $g^{CU}_{ss'}$ cannot be farisghtedly deterred and (ii) using part (i) of Definition 3, any deviation from $g^{CU}_{ss'}$ by $S \subseteq \{s, s'\}$ is farsightedly deterred. Thus, $G_1 = f(\mathbb{G}) = g^{CU}_{ss'}$ and, in turn, $G_1 = f(G_1)$ given that $f(G_1)$ is non-empty and $f(G_1) \subseteq G_1$ (Chwe 1994). Hence, $g^{CU}_{ss'}$ is the LCS. ■

Proof of Proposition 4.

We will establish that $g^{PTA-PL}_{ss',lss'} = f(\mathbb{G})$ and, hence, $g^{PTA-PL}_{ss',lss'}$ is the LCS. To begin, the following two observations establish that $g^{PTA-PL}_{ss',lss'} \in f(\mathbb{G})$: (i) equation (24) says the small countries prefer $g^{PTA-PL}_{ss',lss'}$ over any other network and that (ii) any deviation by country l from $g^{PTA-PL}_{ss',lss'}$ results in either $g = g^{FTA}_{ss'}$ or $g = g^{CU}_{ss'}$ which leaves l worse off via equation

(25). Thus, part (i) of Definition 3 implies any deviation is farsightedly deterred. Thus, $g^{PTA-PL}_{ss',lss'} \in f(\mathbb{G})$.

Two observations establish that $g \notin f(\mathbb{G})$ for $g \neq g^{PTA-PL}_{ss',lss'}$. First, the deviation by $S = \{s, s'\}$ from any $g \notin \{g^{PTA-PL}_{ss',lss'}, g^{CU}_{ss'}, g^{FTA-PL}_{ss',ls}, g^{FTA-PL}_{ss',ls'}\}$ to $g^{CU}_{ss'}$ cannot be farsightedly deterred. To see this, note that l has no deviation from $g^{CU}_{ss'}$. And, equation (24) says any subsequent deviation by s (or s') that ultimately leaves it better off than at g will also leave s' (or s) better off.

Second, deviations to $g^{CU-PL}_{ss',lss'}$ cannot be farsightedly deterred from $g^{CU}_{ss'}$ by $S = \{l, s, s'\}$ or from $g \in \{g^{FTA-PL}_{ss',ls}, g^{FTA-PL}_{ss',ls'}\}$ by the plurilateral non-member. Equations (24)-(25) say the deviating countries have a higher payoff at $g^{CU-PL}_{ss',lss'}$. And, the logic from the first paragraph establishing $g^{PTA-PL}_{ss',lss'} = f(\mathbb{G})$ implies such deviations cannot be farsightedly deterred.

Thus, we have established that $G_1 = g^{PTA-PL}_{ss',lss'} = f(\mathbb{G})$. In turn, $G_1 = f(\mathbb{G}_1)$ given that $f(G_1)$ is non-empty and $f(G_1) \subseteq G_1$ (Chwe 1994). Hence, $g^{PTA-PL}_{ss',lss'}$ is the LCS. ■

Proof of Proposition 5.

Note that equations (27)-(28) say that (i) s prefers $g \in \{g^{FTA}_{ll'}, g^{CU}_{ll'}, g_{\emptyset}\}$ over any other network g' and that (ii) l prefers both $g^{CU}_{ll'}$ and $g^{FTA}_{ll'}$ over $g_{\emptyset}$ but prefers $g^{CU}_{ll'}$ over $g^{FTA}_{ll'}$. Thus, given s is a member of some agreement in any $g \notin \{g^{FTA}_{ll'}, g^{CU}_{ll'}, g_{\emptyset}\}$, s has a deviation from any such g to some $g' \in \{g^{FTA}_{ll'}, g^{CU}_{ll'}, g_{\emptyset}\}$ that cannot be farsightedly deterred because (i) it prefers g' over g and (ii) the only subsequent deviation that could eventuate involves l and l' to $g'' \in \{g^{FTA}_{ll'}, g^{CU}_{ll'}\}$ but s still prefers g'' over g. Moreover, these observations also imply the deviation by $S = \{l, l'\}$ from $g \in \{g_{\emptyset}, g^{FTA}_{ll'}\}$ to $g^{CU}_{ll'}$ cannot be farsightedly deterred. Hence, $G_1 = f(\mathbb{G}) \subseteq \{g^{CU}_{ll'}\}$.

Finally, $G_1 = f(\mathbb{G}) = g^{CU}_{ll'}$. This follows because part (i) of Definition 3 implies any deviation from $g^{CU}_{ll'}$ involving s and/or s' is farsightedly deterred and also because l has no unilateral deviation from $g^{CU}_{ll'}$. In turn, $G_1 = f(G_1) = g^{CU}_{ll'}$ given that $f(G_1)$ is non-empty and $f(G_1) \subseteq G_1$ (Chwe 1994). Hence, $g^{CU}_{ll'}$ is the LCS. ■

Proof of Proposition 6.

Equations (27)-(30) imply that any deviation away from $g^{CU-PL}_{ll',sll'}$ by some $S \subseteq \{s, l, l'\}$ to some network g' is farsightedly deterred because some $i \in S$ prefers $g^{CU-PL}_{ll',sll'}$ over g'. Thus, $g^{CU-PL}_{ll',sll'} \in G_1 = f(\mathbb{G})$.

Using the logic from the proof of Proposition 5, equations (27)-(30) imply $g \notin G_1 = f(\mathbb{G})$ for any network g that involves PTAs and in which s is the member of at least one PTA. Moreover, the only networks s prefers over $g^{CU-PL}_{ll',sll'}$ involve a single FTA between l and l' but, for any given plurilateral agreements, l and l' prefer forming a CU rather than an FTA and, conditional on s not forming any PTAs, l and l' prefer $g^{CU-PL}_{ll',sll'}$ over any other

network. Thus, starting from $g^{CU-PL}_{ll',sll'}$, l and/or l' will not participate in any deviations that lead to a network involving a single FTA between l and l', including following the only unilateral deviation available to s which is from $g^{CU-PL}_{ll',sll'}$ to $g^{CU}_{ll'}$. In turn, there is a deviation from $g \in \{g^{CU}_{ll'}, g_{\emptyset}, g^{PL}_{\cdot}\}$ to $g^{CU-PL}_{ll',sll'}$ by some $S \subseteq \{s, l, l'\}$ that cannot be farsightedly deterred. Similarly, the deviations by $S = \{l, l'\}$ from $g^{FTA-PL}_{ll',\cdot}$ to $g^{CU-PL}_{ll',sll'}$ and from $g^{FTA}_{ll'}$ to $g^{CU}_{ll'}$ cannot be farsightedly deterred. Thus, $g \notin G_1 = f(\mathbb{G})$ for any $g \neq g^{CU-PL}_{ll',sll'}$.

We have thus established $G_1 = f(\mathbb{G}) = g^{CU-PL}_{ll',sll'}$. In turn, $G_1 = f(G_1) = g^{CU-PL}_{ll',sll'}$ given that $f(G_1)$ is non-empty and $f(G_1) \subseteq G_1$ (Chwe 1994). Hence, $g^{CU-PL}_{ll',sll'}$ is the LCS. ■

D Numerical and Computational Details

D.1 Optimal Tariffs

While Appendix Table A.1 provides closed-form expressions for optimal tariffs in the homogeneous sector, we only obtain closed-form solutions for optimal tariffs in the differentiated goods sector for the special case of $\sigma = 2$. In general, for a given trade network and vector of paramaters (α, σ, N), country i's optimal tariffs in the differentiated sector are determined via maximizing its welfare $W_i(\mathbf{t}, \boldsymbol{\tau})$, where $\mathbf{t} = (\mathbf{t}_i, \mathbf{t}_j, \mathbf{t}_k)$ and $\boldsymbol{\tau} = (\boldsymbol{\tau}_i, \boldsymbol{\tau}_j, \boldsymbol{\tau}_k)$, by choosing its vectors of homogeneous sector tariffs $\mathbf{t}_i$ and differentiated sector tariffs $\boldsymbol{\tau}_i$ subject to the MFN constraint and the Article XXIV constraints that it must impose zero tariffs on PTA partners and not raise tariffs on PTA non-members. To do so, we use *Mathematica 12.3* to solve the first-order conditions using `FindRoot`. To ensure convergence of the residual and the solution to approximately 10^{-7}, we set `AccuracyGoal = 7` and `PrecisionGoal = 7`.

To facilitate computation of the LCS across the (α, σ)-paramater space, we discretize the parameter space by considering $\alpha \in [1.5, 6.0]$ with a step size 0.01 and $\sigma \in [1.05, 6.0]$ with a step size 0.01. This yields 451 parameter vectors for the homogeneous sector and $496 \times 451 = 223,696$ parameter vectors for the differentiated sector. We calculate the optimal tariffs and the associated welfare levels for each country across all trade agreement networks for each possible parameter vector.

D.2 Largest Consistent Set

Using the welfare levels evaluated at optimal tariffs from Appendix D.1 and code provided to us by Michael Chwe to compute the LCS in Chwe (1994), we create an algorithm in *Python 3.11.3* to compute the LCS across our parameter space. For each parameter vector, the algorithm:

- Loads welfare values for each country in every trade agreement network $g \in \mathbb{G}$.
- Loads matrices that indicate the (feasible) deviations for each coalition $S \subseteq \{i, j, k\}$.
- Computes direct and indirect dominance relationships between each pair of trade agreement networks $g, g' \in \mathbb{G}$.
- Iterates using the function $f(G)$ for $G \subseteq \mathbb{G}$ as described in Section 3.2 of the main text to find the fixed point of $f(G)$ which is the LCS.

D.3 Replication

A supporting GitHub repository contains all Mathematica and Python files for the replication of the results, along with a detailed README file and requirements file. The repository is available at https://github.com/Lchochua/CLW-paper-supplement.