

THE WORLD TRADE ORGANIZATION AND THE FUTURE OF MULTILATERALISM^a

Richard Baldwin^b

^aAn edited version of this will appear in the Journal of Economic Perspectives in 2016. This paper draws on a several of my unpublished policy essays that have been posted in the CEPR policy discussion papers, Policy Insights (Baldwin 2010, 2011, 2012), and an unpublished paper I wrote for the OECD, Baldwin (2014)..

^b Professor of International Economics, Graduate Institute of International and Development Studies, Geneva.

© The Author.

All rights reserved. No part of this publication may be reproduced without the permission of the author.

Centre for Trade and Economic Integration (CTEI)

The Centre for Trade and Economic Integration fosters world-class multidisciplinary scholarship aimed at developing solutions to problems facing the international trade system and economic integration more generally. It works in association with public sector and private sector actors, giving special prominence to Geneva-based International Organisations such as the WTO and UNCTAD. The Centre also bridges gaps between the scholarly and policymaking communities through outreach and training activities in Geneva.

www.graduateinstitute.ch/ctei

When the General Agreement on Tariffs and Trade was signed by 23 nations in 1947, the goal was to establish a rules-based world trading system, and to facilitate mutually advantageous trade liberalization. As the GATT evolved over time and morphed into the World Trade Organization in 1993, both goals have been achieved. The WTO presides over a rule-based trading system based on norms that are almost universally accepted and respected by its 163 members. Tariffs today are below 5 percent on most trade, and zero for a very large share of imports.

Despite its manifest success, the WTO is widely regarded as suffering from a deep malaise. The main reason is that the latest WTO negotiation, the Doha Round, has staggered between failures, flops and false dawns since it was launched in 2001. But the Doha logjam has not inhibited tariff liberalization – far from it. During the last 15 years, most WTO members have massively lowered barriers to trade, investment and services bilaterally, regionally, and unilaterally – indeed, everywhere except in the WTO. The massive tariff cutting that has taken place around the world, shown in Table 1, has been at least as great as in the previous successful WTO rounds. Moreover, the Doha gridlock has also not dampened nations’ interest in the WTO; 20 nations, including China and Russia, have joined since 2001.

Table 1: Tariff cutting despite the Doha deadlock.

	Percentage points		Change from 2001 to 2012	
	2001	2012	Percentage point difference	Percentage cut
South Asia	22	13	-9	-41%
Middle East & North Africa (developing only)	19	12	-7	-38%
Sub-Saharan Africa (developing only)	14	11	-3	-19%
Latin America & Caribbean (developing only)	11	8	-4	-32%
East Asia & Pacific (developing only)	11	8	-3	-31%
World	10	7	-3	-30%
Europe & Central Asia (developing only)	8	4	-4	-49%
European Union	4	1	-2	-63%

Source: World Bank online database.

Note: Figures are for tariff rates (%), applied, simple mean, all products.

This paper begins by sketching the historical context of the original GATT agreement. It then discusses how the rules and principles behind the GATT rounds combined to create a juggernaut of political economy momentum in which nations kept joining the GATT and tariffs kept falling.

The paper then turns to the current woes of the WTO and why its magic seems to have failed in the Doha Round. Two major sets of reasons emerge in this discussion. First, the last GATT negotiations, the Uruguay Round, sought to generate additional momentum for free trade through broadening its focus, both in terms of more countries joining and in terms of additional areas that would be covered by the agreement. However, these steps toward broadening also required altering some of the historical rules and principles that had generated momentum toward free trade. The changes altered and may even have ended the political economy momentum of the WTO. Second, the rules and procedures of the WTO were designed for a global economy in which made-here-sold-there goods moved across national borders. But the rapid rising of offshoring from high-technology nations to low-wage nations has

created a new type of international commerce. In essence, the flows of goods, services, investment, training, and know-how that used to move inside advanced-nation factories have now become part of international commerce. For this sort of offshoring-linked international commerce, the trade rules that matter are less about tariffs and more about protection of investments and intellectual property, along with legal and regulatory steps to assure that the two-way flows of goods, services, investment, and people will not be impeded.

It's possible to imagine a hypothetical WTO which incorporates these rules. But in practice, the rules are being written in a series of regional, and megaregional agreements like the Trans-Pacific Partnership (TPP) and Transatlantic Trade and Investment Partnership (TTIP) between the United States and the European Union. The most likely outcome for the future governance of international trade is a two-pillar structure in which the WTO continues to govern with its 1994-era rules, while the new rules for international production networks, or "global value chains", are set by a decentralized process of sometimes overlapping and inconsistent megaregional agreements.

The Historical Context for the Principles of GATT

The GATT was launched in unusual times. The 'demand' for trade liberalization was great as tariffs were still high from the Smoot-Hawley tariff and retaliations in the 1930s. The 'supply' of trade liberalization was, in general terms, equally great as leaders of the largest trading nations wanted to avoid the protectionist mistakes of the 1920s and 1930s. The demand and supply of liberalization were also powerfully driven by the outbreak of the Cold War; world trade integration became a geostrategic issue as well as a commercial issue.

The GATT's design was heavily influenced by lessons drawn from historical trade liberalization efforts (Irwin et al 2008). Pre-WWI globalization had few international organizations, supported instead by Pax Britannica. During WWII, the US became the global leader and it wanted postwar globalization to be based on international institutions. The US Congress, however, which controls US trade policy, refused to bind its hands in a new international organization. Instead, trade would be buttressed by a 'general agreement' but no formal organization like the IMF. The Agreement was based on several principles.

One General and Five Specific Principles

There is no definitive list of principles in the GATT and WTO, and authors differ on the matter (for a detailed account, see Hoekman and Kostecki 2009). However, it is useful to think of one general and five specific principles. The general principle - what might be called the constitutional principle - is that the world trade system should be rules-based, not results-based. The GATT, and now the WTO, focuses on the design, implementation, updating and enforcement of procedures, rules and guidelines, rather than on seeking to agree upon, for example, the volume of exports or market shares. This overarching constitutional principle is implemented with five specific principles.

1) *Non-discrimination*. This rule has two aspects: non-discrimination at the border, and non-discrimination behind the border. Nondiscrimination at the border, which is called "most favored nation" in WTO circles, means that any tariff which is applied should be applied equally to all WTO members. Many exceptions are allowed (for example, free trade agreements), but these are controlled by explicit conditions. The other aspect of non-discrimination is called "national treatment," which is the rule that within each country, taxes and regulations should be applied evenly to domestic and imported goods.

2) *Transparency*. Liberalizing trade and reducing conflicts over trade is easier when the actual policies are transparent to all by having been made public.

3) *Reciprocity*. Nations that remove barriers to imports can expect other nations to reciprocate. Again, exceptions are made, with the most notable example being that, during the GATT era, developing nations benefitted from other nation's market opening due to the most-favored-nation provisions, but they were allowed not to reduce their own tariffs. Reciprocity also applies to retaliation. When a nation engages in a practice or policy that undoes the gain another member had from a previous agreement, the aggrieved nation has the right to reciprocate—that is, to retaliate.

4) *Flexibility, or "safety valves."* The founders of the GATT knew that members would occasionally be subject to irresistible domestic pressure to impose trade barriers. Rather than threatening implausibly dire consequences for such actions, the GATT allows some exceptions in which nations can at times impose trade barriers, but seeks to discipline them with various strictures and requirements for compensation.

5) *Consensus decision-making*. Like the other principles, this one has exceptions, but most WTO decisions are by consensus.

As the next section explains, interactions among these principles generated a political economy momentum that drove trade liberalization. As the following section explains, changes made in the 1990s help to explain why the momentum has ground to a halt.

A Tariff-Cutting Juggernaut

GATT is widely viewed as having facilitated the reduction of tariffs—at least in the developed nation. Systematic data on tariffs for a broad range of nations is available only from the 1980s, but a cruder measure called the “effective tariff rate”—that is, tariff revenue divided by the value of imports—has been collected back to the beginning of GATT by Clements and Williamson (2004). An obvious problem with the effective tariff rate measure is that really high tariffs result in very low imports and so tend to get little weight in the average. In addition, the effective tariff rates for individual nations can be very noisy over periods of only a few years, because they reflect both changes in tariff rates and changes in patterns of imports. Despite these well-known problems, effective rates give a reasonable general idea of tariff cutting patterns under the GATT.

Two salient facts that emerge from the top panel of **Figure 1** are that low-income nations have always had higher tariffs and developed nations reduced their tariffs steadily while poor nations only starting doing so in the 1980s. In looking at the figures, remember that up-and-down fluctuations in effective tariffs during periods of a few years maybe due to shifts in the price or quantities of specific imports, along with changes in tariff policies.

Four Phases of Trade Liberalization under the GATT: 1950-1994

Figure 1 presents averages of effective tariff rates across selected developed and low-income nations. Four phases are distinguished based on the timing of GATT Rounds.

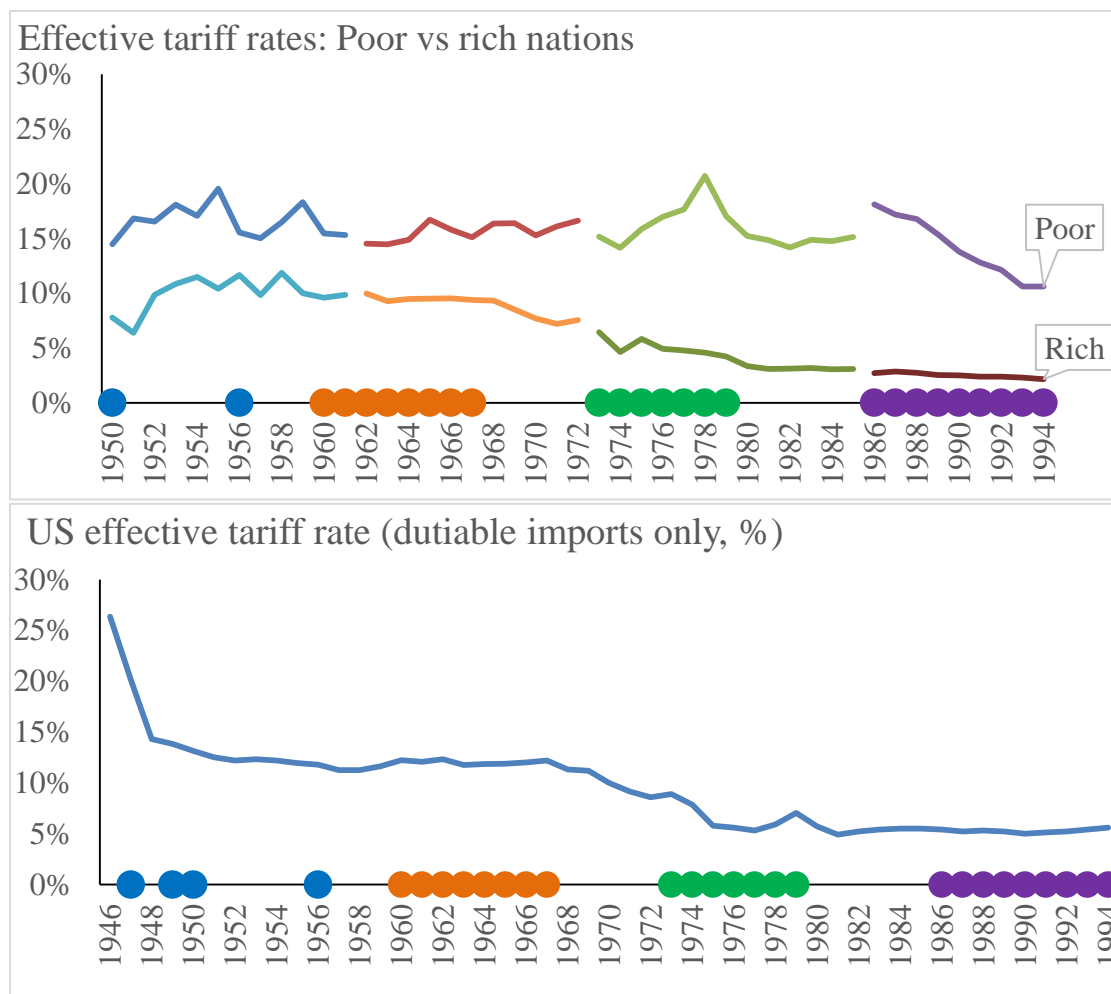


Figure 1: Effective tariff rates, developed and low-income nations, 1950 - 1994.

Source: (top panel) Clemens and Williamson (2004) data with author's elaboration; (bottom panel) US Historical Data with authors elaborations.

Note: Developed nations comprise the European Union nations plus Switzerland, Norway, Japan, and Australia. Low-income nations comprise Argentina, Brazil, China, Egypt, Indonesia, India, Kenya, Korea, Mexico, Malaysia, Nigeria, Pakistan, Philippines, Thailand and Turkey. The dots at the bottom indicate a GATT/WTO Round is underway. The bottom chart shows the effective tariff rate for the United States.

The first phase, up until 1960, began with a substantial wave of tariff-cutting in the 1947 inaugural Geneva Round. However, the other early Rounds were not focused on tariff-cutting. Instead, they considered details of rules and accessions such as those of Germany (1951) and Japan (1955). Moreover, tariffs were not the main trade hindrance to international trade in the 1950s. Instead, restrictions remaining from wartime, along with state trading and inconvertible currencies, were the binding constraints. For example, European currency only became generally convertible in 1958.

The second phase from 1960 up to 1972 was triggered by European regional trade liberalization (Kenen 1989; Preeg 1970). For example, the Dillion Round (1960-61) dealt with the tariff concessions EU members had to make to other GATT members in compensation for the formation of their customs union.³ The Kennedy Round (1963-1967) was, in part, an effort by the US, Japan and other large exporters to redress the trade diversion arising from this

³ The EU consisted of France, Italy, Germany, Netherlands, Belgium and Luxembourg at the time; the common external tariff chosen required the Benelux nations to raise bound rates, France and Italy to lower them and Germany to change very little. Under GATT rules, other GATT members could demand compensation for any of the tariff rises.

customs union. The decline in tariff rates in developed countries after about 1967 was in part due to GATT, but also to non-GATT steps like elimination of tariffs across much of Europe, and the US-Canada Auto Pact of 1965 which eliminated tariffs on bilateral auto trade. Note how regionalism and multilateralism advanced hand-in-hand in this phase. By contrast, tariffs in low-income nations did not fall since GATT rules excused them from reciprocally cutting their tariff in GATT talks (this is an exception to the non-discrimination principle called “Special and Differential Treatment” for developing nation members).

The third phase of trade liberalization started around 1973, and again multilateralism and regionalism advanced together. The GATT’s Tokyo Round talks were launched the same year that the European Union enlarged (Britain, Ireland and Denmark joined) and signed bilateral free trade agreements with most other West European nations. The 1970s are a period when the “effective tariff” measure can be deceiving. It looks as if substantial tariff cutting happened in developed nations, although no GATT-required cuts were implemented until the Round finished in 1979. The illusion arises from the 1970s price hikes that raised the import shares for oil; as developed nations had low or zero tariffs on imported oil, the relative price change looks like a cut in the average tariff. The US numbers in the bottom panel of **Figure 1**, which strip out duty-free imports (which includes petroleum in this case), show no tariff reductions at this time.

The fourth wave of multilateral and regional trade talks arose in mid-1980s. In 1986, for example, GATT members launched the Uruguay Round, the United States and Canada started talks about a bilateral free-trade agreement, and the European Union enlarged to include Spain and Portugal while launching its Single Market Program that eliminated a vast range of non-tariff barriers to cross-border movements of goods, services, and workers. Effective tariffs fall gently in developed nations, probably mostly due to regional rather than multilateral liberalization. For example, at the time the Uruguay Round was launched in 1986, about 40 percent of global trade took place inside free-trade areas, with about half of that within the European Union. The really original element in this fourth phase was the rapid tariff cutting by developing nations—but this they did outside the GATT and was driven by changes in their attitudes towards high tariffs on industrial goods (more on this change below). Developing nations also signed many regional trade agreements (RTAs), like Mercosur in South America and the South African Customs Union. These had some effect on tariffs, but many developing nations lowered their multilateral tariffs at the same time as they cut tariff with their partners in free-trade agreements (Estevadeordal et al 2008).

As this sketch of the four phases reveals, the momentum toward cutting tariffs includes both multilateral and RTAs. Thus, the underlying question is what generates this kind of political economy momentum.

The Juggernaut Dynamics of Tariff Cutting

Tariffs, like most economic policies, are the outcome of a political economy process. To explain why governments lower tariffs they previously found politically optimal to impose, the literature points to the role of trade agreements (Bagwell and Staiger 2004). The basic approach models the process as a one-time switch from a non-cooperative outcome to a cooperative outcome facilitated by a trade agreement. This helps explain the initial drop in American tariffs at the start of the GATT (see **Figure 1**), but a switch from one form of equilibrium to another leaves out most of the richness of how the GATT fostered multiple

forms of tariff-cutting over successive Rounds.⁴ In addition, it does not explain why developing nations acted outside the GATT to cut their tariffs starting in the mid-1980s.

More elaborate approaches to the political economy of tariff-cutting draw on the intuitive, if informal, two-level game approach of Putnam (1988) as formalized by Grossman and Helpman (1995). This approach argues that governments negotiate both with special interest groups within their nation and other governments internationally. The discussion here is organized around a version of the two-level-game approach that I introduced in 1994, which I called the “juggernaut effect.” It is easiest to explain in the historical context.

Before the GATT, exporters had only a very indirect interest in their nation’s import tariffs. But under the GATT reciprocity principle, foreign tariff levels became linked to domestic tariff levels. Of course this only held for developed nations who followed the reciprocity principle. In a way, the GATT’s success was not due to the international deal itself. It was due to the way the principles behind the international deal altered domestic political realities in developed-nation members. Also, remember that developing nation governments were excused from reciprocity by the Special and Differential Treatment rule, and thus faced the same array of domestic pro- and anti-tariff special interests before, during, and after each GATT round. In theory and practice, this meant that they did not lower tariffs that they previously found optimal to impose.

In the juggernaut story, the first round of tariff cuts creates political economy momentum. As tariffs drop, pro-tariff import-competing firms face additional international competition. Many of them shrink, become less profitable, and even go out of business. Conversely, foreign tariff-cutting boosts exporters. They expand and become more profitable. In this way, a one-off tariff cut weakens protectionist forces and strengthens liberalization forces from a political economy perspective. A few years down the road, when another multilateral GATT round is launched, the altered political economy power of importers and exporters comes into play. As before, exporters have an incentive to fight for domestic tariff cuts due to the reciprocity principle, and import-competing firms have an incentive to fight against them. But since the anti-liberalization camp is systematically weaker and the pro-liberalization camp is systematically stronger than during the last round, all the governments playing reciprocally find it politically optimal to cut tariffs again. As these fresh tariff cuts are phased in, the exit of import-competing firms and entry of exporters again reshapes the political landscape inside each participating nation, and the cycle restarts. The juggernaut rolls forward.

This dynamic also suggests an explanation for why multilateral and regional tariff cutting progressed in tandem. Once the original tariff cuts weaken protectionists and strengthen liberalizers, governments find it optimal to lower tariffs both multilaterally and regionally.

A related political economy dynamic is that RTAs can kick-start multilateral trade liberalization. For example, a quick eyeballing of **Figure 1**, especially the US effective tariff rates, suggests that the juggernaut had run low on momentum by the end of the 1950s. However, when the countries of Western Europe began cutting their intra-European tariffs from 1959, the resulting tariff discrimination aroused the concerns of exporters in the US, Japan and Canada.⁵ At that time North America and Japan both sent roughly one-third of their exports to Europe, and their firms feared losing these markets to European firms who enjoyed zero tariffs inside the customs union. In addition, the freeing of trade within Europe tended to strengthen European exporters and weaken importers. A similar thing may have happened when the

⁴ Lockwood and Zissimos (2004) propose a Bagwell-Staiger-type model where tariff cutting is gradual due to retaliation limits that are, in their model, essential for supporting reciprocal tariff cutting.

⁵ The GATT-EEC link is explicit in President Kennedy’s “Special Message to the Congress on Foreign Trade Policy” (January 25, 1962, <http://www.presidency.ucsb.edu/ws/?pid=8688>).

European customs union was enlarged in 1973 – the same year that the Tokyo Round talks started.

Avoiding Backsliding: Binding plus Allowing Retaliation

The GATT had other mechanisms to keep this gradual, mutually advantageous tariff cutting on track. After all, the “juggernaut” rules of political economy momentum can work in reverse—as it had in the 1930s. Thus, the GATT process included a set of rules designed to make political reversals difficult for individual members. One rule was the principle that a nation’s past tariff cuts were “bound” in the sense that previously agreed tariff levels were not open to further negotiation. Moreover, a nation’s partners could retaliate against any violation of such tariff “bindings”. The effect was to ensure that each nation’s exporters would be punished for any backsliding by its own government. This gave them an incentive to push their government to respect the bindings. Notice that this design element did not depend on the nation’s own government – it was enforced by the risk that foreign governments would retaliate by raising tariffs.

Three Escape Hatches

How could the many countries of the GATT reach agreements while working on a consensus principle? One answer is that some escape hatches were historically allowed, which made it easier for members to agree the tariff cuts in the first place.

As one example of an escape hatch, a variety of GATT practices on “Special and Differential Treatment” meant that developing nations were not subject to GATT disciplines. They were exempted from an expectation of reciprocally cutting their tariffs, and they could mostly ignore any GATT rules with which they didn’t agree. In short, the low-income nations that were part of GATT could typically follow a policy of “don’t obey, don’t object.” However, being excused from reciprocity did not mean the developing nations were indifferent to the GATT’s success. The GATT’s most-favored-nation principle meant that the tariff cuts agreed among the developed nations were automatically extended to developing nation exporters. They were free-riders who liked the ride. The developed countries were mostly happy to allow this free riding since developing nation markets were, at the time, rather insignificant.

A second kind of escape hatch emerged in the 1960s and 1970s during the Tokyo Rounds, in which negotiations on trade rules were undertaken by the so-called “Codes” approach. This meant that each set of rules agreed upon was adopted in the form of a code, which would be binding only for those members that voluntarily signed them—which in practice typically meant the developed nations. For example, during the Tokyo Round a number of issues beyond tariffs were put on the agenda using the “codes” approach – many of these involving new forms of protection that had arisen in the 1960s and 1970s to offset competitive effects of earlier tariff cuts (Baldwin 2009, 1970). However, the principle of non-discrimination meant that countries which did agree to these codes were (mostly) obliged to extend the rules to all GATT members, even those that did not sign the codes.

A third escape hatch arose because the GATT dispute settlement system wasn’t strong enough to enforce compliance. Disputes were brought before a panel whose rulings were reviewed by a group of members that included the disputing parties. According to the consensus principle, the Panel ruling was only accepted if all parties agreed. For example, in 1959 the European Free Trade Association (EFTA) nations wanted free trade among

themselves, but only on industrial goods. In 1965, the United States and Canada wanted to liberalize bilateral trade in the auto sector. When GATT panels were formed to investigate the ‘GATT-legality’ of these limited free trade agreements, the US and EFTA nations blocked the panel from reaching a conclusion.

Of course, if GATT members had extremely diverse preferences, escape hatches like blocking the dispute resolution process could have become a main exit, thus rendering the rules useless. But instead, the combination of a dispute procedure with an escape hatch facilitated agreements by allowing GATT members to be satisfied with ‘constructively ambiguous’ wording. The GATT’s quasi-legal dispute mechanism with escape hatch could be relied upon to settle disputes or at least to help frame future negotiations aimed at clarify ambiguities if and when such clarification proved important.

Causality

The story as told hereto has been of the GATT causing tariff cuts, but how do we know that it was not a third effect causing both GATT membership and tariff cutting? The prima facie evidence is clear, even if the econometrics has not been done due to the lack of high-quality historical tariff data. Two types of tariffs were not subject to the juggernaut ‘treatment’—all developing-nation tariffs, and agricultural tariffs of all GATT members. Neither set of tariffs fell during the GATT days; agricultural tariffs since they were not on the negotiating table, and developing-nation tariffs since they were excused from reciprocal cuts. This suggest that ‘controlling’ for country characteristics (like propensity to cut), the juggernaut treatment only worked on the tariffs it was applied to.

Refueling the Juggernaut, But Closing the Escape Hatches

By the 1970s, tariffs in the developed nations were already fairly low - at least on the products on which they had been willing to negotiate. Agriculture and labor-intensive industrial goods like clothing had been explicitly taken off the bargaining table when the agendas were set for the earlier GATT rounds. In this way, the GATT liberalization resulted mostly in tariff cutting in areas that most interested developed nation exporters, basically industrial goods. Developing nation exporters, who didn’t have any “skin in the game” due to Special and Differential Treatment, were often disappointed in the lack of liberalization of agriculture and labor intensive manufactures.

To refuel the trade liberalization juggernaut, the developed nations that had mostly driven the GATT process decided to broaden the agenda. The process started during the Tokyo Round with the “codes” approach to including nontariff issues in the negotiations. Then with the Uruguay Round starting in 1986, new areas of interest to exporters in developed nations were put on the negotiating table, notably intellectual property issues, restrictions on foreign investment and exported services issues. These came to be known as TRIPs (Trade-Related Intellectual Property), TRIMs (Trade-Related Investment Measures), and services, respectively. Additionally, two sectors still marked by high tariffs - agriculture and clothing - were put on the table to fuel the interest of agriculture exporters and low-wage exporters. This constellation of new issues was viewed as balancing the refueling of the juggernaut along North-South lines. Northern exporters were to gain from new rules and new market access in TRIPs, TRIMs and services, while Southern exporters were to gain from freer trade in food and clothing.

However, the dynamics of the negotiations and the increasing importance of emerging market economies meant that together with broadening the agenda, some of the earlier escape hatches were closed up. For example, industrial nations' domestic laws already assured intellectual property protection for foreigners, so the expected gains for intellectual-property exporters from developed countries would come primarily from getting developing nations to adopt the standards of developed countries on patents, copyrights and the like. During the Uruguay Round, developed countries feared that their opening of agriculture and textile markets would be pocketed by developing nations, but the export gains they hoped for from new disciplines on TRIPs, TRIMs and services would be picked apart. A voluntary codes approach would simply not do for a deal balanced in this way. The developing nations most likely to be affected would be those most likely to opt out.

As a result, the Uruguay Round ended up including a feature called the Single Undertaking. All members, developed and developing alike – even those that had not participated actively in the negotiations – were obliged to accept all the Uruguay Round agreements as one package. The basic outlines of the package-deal approach had been discussed in December 1991 (Croome 1996). Nevertheless, it clearly came as a surprise to many developing country members, especially those that did not follow the Uruguay Round through its eight years of twists and turns.

In addition, because the new areas involved considerable ambiguity and newness, members participating in the Uruguay Round negotiations decided it was necessary to greatly reduce the wiggle room in the dispute procedure. Both North and South feared that exporters' gains in the new areas might be offset by murky forms of protection or slippery national interpretations of the rules. For example, many governments in emerging economies were concerned that the United States was prone to taking unilateral action against whatever the US government considered to be an unfair trade practice under Section 301 of the Trade Act of 1974 (Keohane and Nye 2001). The Uruguay agreement eliminated the possibility of blocking the initiation of a dispute resolution or the adoption of a panel ruling, and applied this to all the areas in the Single Undertaking. The new adjudication procedure welded shut the earlier escape hatch.

Win-Win Multilateral Cooperation

From its start in 1946 until it was superseded by the World Trade Organization in 1995, the GATT promoted win-win multilateral cooperation by setting up what Douglas North would call an “institution” – constraints that guide political and economic interactions consisting of formal rules and informal restraints. The principles of the GATT fostered a self-enforcing pattern of cooperation and success. As the GATT's liberalization process started working its magic, exports of manufactured goods boomed – growing twice as fast as manufacturing output from the late 1960s until just before the collapse of trade in 2009. Booming trade and incomes strengthened GATT members' belief that following the code of conduct was good policy. As nations and interest groups came to expect that the rules would be respected, they adopted behaviors that conformed to the rules thus making observance easier.

The Woes of the WTO

A performance review of the WTO would produce an unbalanced report card. Little progress has been made on the trade liberalization front for almost two decades, since a handful of agreements in 1997. The Doha Round that started in 2001 is stalled. Of the WTO's

functions, only the dispute settlement mechanism would receive a high performance score. Why did the GATT trade liberalization magic stop working for the WTO? I consider both external and internal reasons, and then consider the implications for multilateral and regional trade talks.

- *External Sources*

The most commonly cited causes of the WTO's difficulties concerns the advanced economies' loss of dominance. This came in two ways. First, as discussed above, the GATT was all about exchanges of market access, so market-size was the coinage of the realm. In the GATT period, the US, EU, Japan and Canada – known as the Quad – dominated on this metric, accounting for 2/3rd of world imports. The rapid growth of emerging economies changed this. Today, the Quad accounts for only half of world imports. Second, the sheer number of developing country members has shifted power in the organization and made talks more difficult. Since the last successful negotiation was launched in 1986, over 70 developing nations have joined, about half of them since the WTO was created. Importantly, this includes China who rejoined in 2001 (having quit two years after joining in 1948), but it does not include the other BRICS who were all founding GATT members.

In theory, more nations does not necessarily hinder tariff cutting: after all, more nations could mean more demand and more supply for better market access. In practice, however, developing countries became active in many more new defensive coalitions (that is, groups interested in preventing better access to their own markets) than new offensive coalitions (groups interested in getting better access to foreign markets) (Patel 2007). The reason is straightforward. The reciprocity principle and small size of most developing markets limited their ability to ask foreigners to open up their markets. Hence, such countries had little to gain from new offensive coalitions. The consensus principle, by contrast, gave developing-nation coalitions a good deal of blocking power, which they used to block efforts to open their most politically sensitive markets.

Regionalism also created challenges. RTAs have always been part of the trade governance landscape. From around 1990, however, they played a very different role as the number of agreements skyrocketed. As all of these involved tariff-cutting that would otherwise have had to be funneled via the WTO, and took up political economic 'capital', the rise of regionalism probably made it harder to conclude the Uruguay Round, and is one reason it is so hard to conclude the Doha Round. Things would be easier if the WTO were the "only game in town".

Many of these new agreements were "deep" in the sense of Lawrence (1996), namely, they included legally binding assurances aimed at making signatories more business-friendly to trade and investment flows from other signatories. This can be seen in WTO data that codifies the content of RTAs (based on seminal work by Horn et al 2010). **Figure 2** (left panel) shows the flow of new agreements and the flow of new deep provisions in them (according to my classification that picks out provisions related to offshoring). At about the same time, an old form of economic integration agreement became very popular, i.e. Bilateral Investment Treaties (**Figure 2**, right panel). Basically, these are concessions of sovereignty undertaken to encourage inward investment. For example, signatories commit to resolve investor-state disputes in a forum based in Washington DC rather than in national courts. In their heyday, hundreds of such deals were signed annually. By the late 1990s, most developing nations had already signed them with their major investment partners, so the number fell off sharply. There are now over 3,000 such agreements in existence.

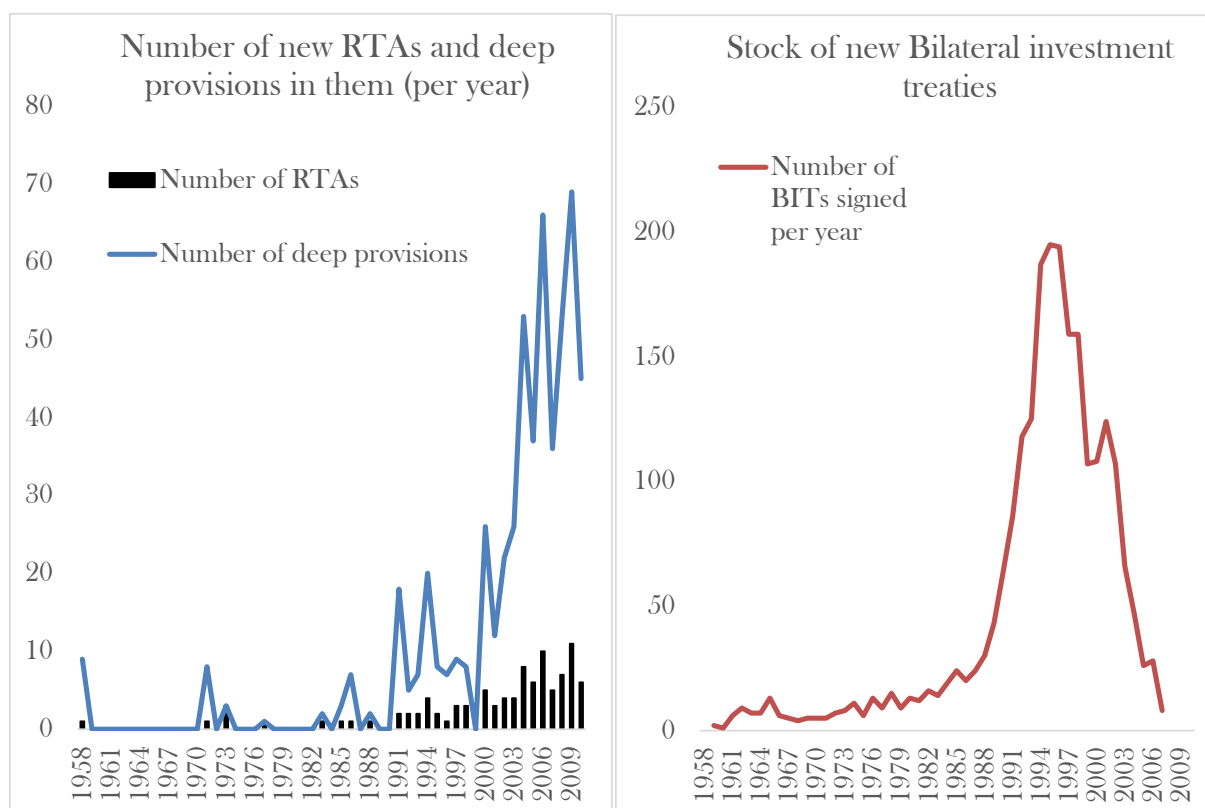


Figure 2: Number of RTAs, deep provisions and BITs.

Source: WTO RTA database (left) and UNCTAD online data (right). The provisions counted as deep are those that constrict nation laws on foreign investment, intellectual property rights, regulatory convergence, short term movement of managers and technicians, and capital flows.

The boom in the investment treaties and deep provisions did not create a direct ‘competitor’ to the WTO. But they provide revealed-preference evidence that many WTO members were looking for disciplines that went far beyond the “shallow” disciplines included in WTO talks. In other words, the ‘demand’ for policy reforms shifted away from the sort of disciplines that the WTO was set up to negotiate. A different challenge came from developing nations’ unilateral tariff cutting.

The rise of offshoring opened a new pathway to industrialization. The old, import-substitution path meant building a supply chain at home in order to become competitive abroad. High tariffs were part of this. The new offshoring-led path involved joining an international production network to become competitive and industrializing by expanding the quantity and range of tasks performed (Baldwin 2011). In this new development model, tariffs hinder rather than help industrialization, so developing nation tariffs started to fall rapidly independently of WTO talks (Figure 3). To maintain flexibility, the developed nations did not ‘bind’ the tariffs in the WTO even when they lowered them on a non-discriminatory basis.

As two-way tariff cutting was the juggernaut’s main ‘fuel’, this unilateralism made multilateral talks less attractive to many developed members whose exporters saw their sales to developed nations boom even as Doha Round staggered from failure to failure. Why fight domestic protectionists at home when foreigners were lowering their tariffs unilaterally?

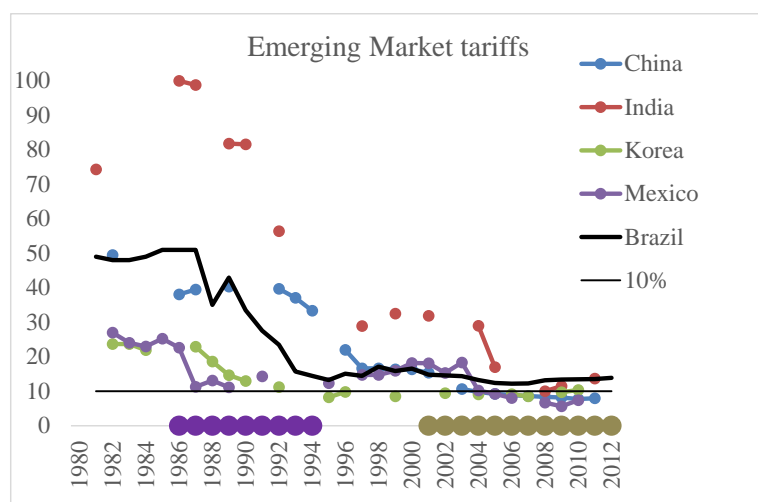


Figure 3: Unilateral tariff cutting by Emerging markets from mid-1980s.

Source: Author's elaboration of World Bank online data.

Note: Simple average of applied MFN tariffs, all products. The dots at the bottom indicate a GATT/WTO Round is underway.

- ***Internal Sources***

These external challenges were magnified by big changes in the way WTO talks were organized as opposed to those under the GATT. To put it bluntly, GATT multilateral negotiations involved the Quad bargaining among themselves over tariff-cuts that they allowed the developing nation members to free ride upon. WTO negotiations, by contrast, require binding tariff cuts and other policy commitments from all but the poorest members. Thus in a political economy sense, the WTO and GATT are quite different international organizations. Specifically, as the Doha Round results would be binding equally on every member unless explicit exceptions were made, the don't-obey-don't-object option that developing nations had under the GATT was cancelled under that WTO (Oyejide 2002). Not surprisingly, they have been far more vocal in the Doha Round than they were in GATT rounds – objecting to provisions that threatened their domestic interests.

- ***Implications for Multilateral Trade Talks***

The impact of these challenges was not immediately apparent. In the years following the 1994 agreement that set up the WTO, multilateral talks worked much as before. A few bits of leftover business, like the 1997 Financial Services Agreement and Information Technology Agreement, were handled in the usual fashion in negotiations led by developed economies. But as the Doha Round got underway, the world discovered that the GATT's juggernaut magic would not work in the WTO. Specifically, the challenges had three momentous implications for the WTO multilateral negotiations.

First, multilateral negotiations under the WTO are more difficult. As explained above, the Single Undertaking principle meant that instead of four veto-players (the Quad, i.e. the US, EU, Japan and Canada) and dozens of free riders as under the GATT, the Doha Round has over a hundred veto-players. Second, business interest in the Doha Round is much less forceful. The agenda for the talks, set in 2001, focused on tariff cutting in industrial goods, and trade distortions in agricultural and service sectors. Industrial trade accounts for 80% of all trade, but business interest was dampened by the fact that tariffs in the Quad nations were already low, and those in the major developing nations had been lowered unilaterally. From a WTO perspective, the exporters of developed nations were now the free-riders on unilateral

tariff cutting by emerging markets. This greatly reduced their interest in lobbying their own governments for a Doha deal.

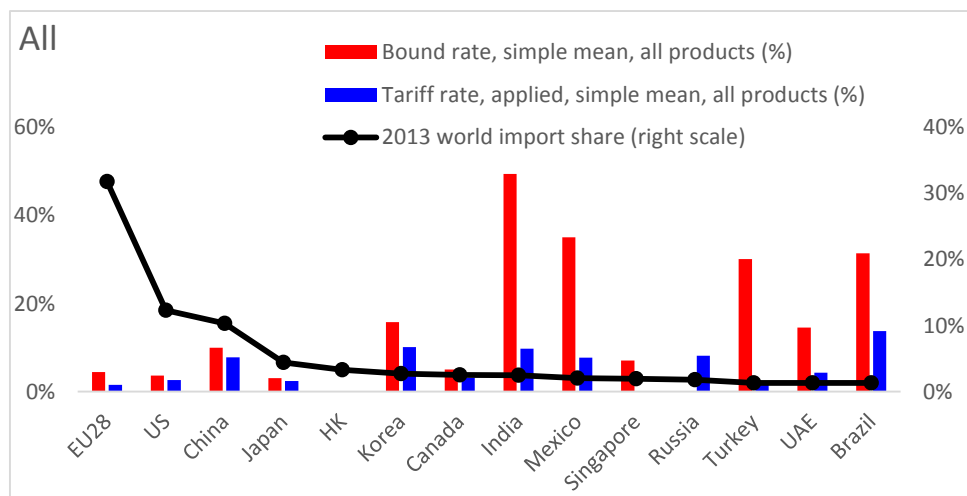


Figure 4: Bound and applied rates for 14 largest importers, 2013.

Source: World Databank for tariffs; WTO online database for imports, with author's elaboration.

Note: these nations account for 80% of world imports in 2013.

One detail of WTO procedures makes unilateral tariff-cutting a major problem for the Doha Round. Following long-standing practice, WTO tariff-cutting talks focus on “bound” tariff rates, not applied rates. For many WTO members, actual applied rates are so much lower than the bound rates that the proposed Doha cuts would only reduce the distance between bound and applied tariffs, without actually lowering the applied rates. **Figure 4** illustrates the issue. Bound and applied rates are shown for the 14 largest importers who together account for 80 percent of world imports. These are the markets to which exporters want access. Five decades of GATT talks had already lowered bound rates in the developed economies to under 3 percentage points on average. In most of the large developing nations, bound rates are quite high, but applied rates are lower. Even in China, the third-largest global market for exports, the applied rate is about 8%. If the developing nations had not lowered their applied rates so much below their bound rates, developed nation exporters would have had something to fight for.

Similarly in agriculture, the biggest protectionists – the EU and Japan – unilaterally lowered distortions for purely domestic reasons. The political power of rich-nation farm lobbies has dropped as farm populations have fallen and awareness has risen about the fact that most farm support goes to wealthy landowners and agri-corporations. The EU broadly switched its agriculture support policies to non-trade distorting forms and basically eliminated export subsidies in major reforms in 2003, 2008 and 2013. Japan still has astronomical tariffs on a handful of products like rice, but it too is shifting unilaterally towards non-trade distorting policies with major reforms in 2003 and 2007. While agriculture trade is hardly free and fair, and the US has increased trade distortions with its 2014 Farm Bill, the mercantilist gain from concluding the Doha Round is clearly lower in 2016 than it was in 2001. Moreover, a number of emerging markets have deployed some of their new found wealth in the form of new trade-distorting agriculture policies. They are, in essence, reacting to exactly the same rural-urban domestic politics that produced agriculture protection in the US, EU and Japan. This has created new opponents to agricultural trade opening.

Third, the rise of offshoring has created a political economy demand and supply for disciplines that underpin international production networks. As these disciplines were not included in the Doha Round's 2001 agenda, and dozens of WTO member have vetoed all moves to expand the agenda, the supply and demand are meeting outside the WTO – mainly

in the deep RTAs and BITs. The point is simple. The rapid rise in production unbundling – sometimes called ‘global value chains’ – meant that the world’s most dynamic trade involves a ‘nexus’ of trade in goods, services, know-how, physical investment, key personnel and financial capital. Many developing nations sought and are still seeking to attract this offshoring activity. Firms in the advanced nations are interested in providing it – as long as they have assurances that host nations would respect their tangible and intangible property rights, and ensure that the necessary flows of goods, services, investment, capital and people would be unimpeded.

These assurances have been provided in dozens of deep bilateral and RTAs and Bilateral Investment Treaties. This ‘spaghetti bowl’ of agreements is clearly not optimal for international production sharing, so a number of so-called megaregionals like the TPP and TTIP have emerged to multilateralise some of the disciplines as a regional level. To understand this, Ethier (1998) presents a model of how development led by foreign direct investment could affect multilateralism. In short, the political economy switched from “my market for yours” to “my factories for your reform”—that is, developing nation trade liberalization and pro-business reforms in exchange for production facilities from developed nations.

The Future of Multilateralism, Regionalism and the WTO

The WTO is a pillar of multilateral economic governance as was the GATT before it. Its prime mission is to establish ‘rules of the road’ and facilitate negotiation of mutually advantageous trade liberalization. In the main, it can claim “mission accomplished”. It oversees a set of near-universal norms for rule-based trade, and it runs a dispute settlement mechanism that routinely arbitrates disputes and issues rulings that are universally followed even though it has no direct enforcement power. Most telling of all, nations ‘vote with their feet’ by joining the WTO even though the requisite reforms typically involve high domestic political costs.

The organization, however, seems frozen in time. The last updating of its rulebook and its last major trade liberalization came in 1994 when Bill Clinton, Gerhard Schroeder, Hashimoto Ryutaro, and Li Peng were in power and the internet did not exist. Likewise the current WTO talks, the Doha Round, are focused entirely on 20th century issues such as tariffs on industrial and agricultural goods, and trade distorting policies in agriculture and services.

While a couple of small agreements have been completed, the Doha Round is in its 15th year and nowhere near done. This 15 year ‘fail trail’, however, has not stalled global trade opening and rule-writing. For 20 years, new rule-writing and new liberalization has proceed apace along three axes – all of them outside of the WTO’s ambit. First, a great deal of tariff-cutting has been done unilaterally by WTO members, especially developing-nation members. Second, new disciplines on international investment – flows that are now intimately entwined with trade in goods and services – have been established by a network of over 3,000 Bilateral Investment Treaties. Third, the new rules and deep disciplines that have underpinned the rapid expansion of offshoring and the internationalization of production have been written into deep RTAs, especially those between advanced and emerging economies.

These observations beg two questions: Is this situation bad? And What is the future of the WTO and multilateralism?

Is bilateralism bad?

Two decades ago, the explosion of bilateral deals shown in **Figure 2** sparked a debate on multilateralism versus regionalism. Authors such as Jagdish Bhagwati decried regionalism as

dangerous (Bhagwati 1993). He pointed to a ‘small-think’ danger – that the inefficiencies of trade diversion would diminish welfare – and a ‘big-think’ danger – that regionalism would block the path to global free trade. As it turned out, global tariff cutting since the rise of regionalism has proceeded as quickly as ever, but outside the WTO (**Figure 1** and **Figure 4**). As a result, the specter of trade diversion never really appeared. Measures based on detailed tariff data show that little of world trade is affected by tariff preference margins of 5 percent or more (WTO 2011). MFN tariffs are zero or very low on most of the world’s large trade flows, so preference margins cannot be high. And where tariffs remain high, the preferences tend not to be granted. That is, bilateral deals exclude such ‘sensitive’ items so no preference is created. In any case, the econometric evidence suggest the trade diversion is not a first-order concern in the world economy (Estevadeordal et al 2008, Acharya et al 2011).

As for the systemic, big-think danger it is hard to know what would have happened if somehow nations had not signed the hundreds of bilaterals that they did. But one thing is clear. The preferential tariff aspect has not blocked the path to global tariff cutting. Virtually all of the developing nation WTO members who engaged in bilateral, discriminatory liberalization have simultaneously been engaged in unilateral, non-discriminatory liberalization.

Importantly, the trade creation/diversion logic only applies to liberalization that is truly discriminatory. Many of the deep RTA provisions concern things where discrimination is impractical. Such disciplines impinge upon corporations, services, capital, and intellectual property. The fact is that it is difficult to write rules that identify the nationality of such things in a way that clever lawyers cannot get around. For example, the Japan-Thailand RTA allows Japanese banks to sell certain financial services in Thailand. But since it is difficult to determine which banks are Japanese, the agreement grants the privilege to any bank registered and regulated in Japan – which makes most large US and EU banks “Japanese” for the purposes of the agreement. This phenomenon of ‘soft preferences’ also arose from the EU’s Single Market program (which is the biggest and deepest of all RTAs). As it turned out, many EU reforms were helpful to non-EU firms even though their nations were not signatories.

Future of Multilateralism

The WTO’s paralysis in the face of frenetic tariff-cutting and rule-writing outside the organization is down to two things. First, the Doha Agenda was set for a world we no longer live in. If it had been concluded in a few years as planned, the juggernaut effect might have worked. But now after the rise of China, the rise of offshoring, and the rise of unilateralism, the negotiating items on the agenda no longer provide a win-win bargain for all. Second, the natural step of expanding the agenda to include some of the disciplines routinely agreed in deep RTAs is blocked by nations who were left aside by the rise of offshoring. They feel that they were promised, in 2001, ‘rebalancing’ that would involve reduced barriers to exports of agricultural and labor-intensive goods. Until they get their rebalancing, they veto an expansion of the agenda.

Since there are important network externalities to be won by moving away from bilateralism and towards multilateralism when it comes to some deep RTA provisions, the WTO’s paralysis led to deals being done elsewhere. The thousands of Bilateral Investment Treaties, for instance, are not that different so network externalities could be realized by melding them together. The emergence of so-called megaregionals like the TPP and TIPP should be thought of as partial multilateralisation of existing deep disciplines by sub-groups of WTO members who are deeply involved in offshoring and global value chains.

The megaregionals, however, are not a good substitute for multilateralization inside the WTO. They will create an international trading system marked by fragmentation (because they are not harmonized among themselves) and exclusion (because emerging trade giants like China and India are not members now and may never be). Whatever the conceptual merits of moving the megaregionals into the WTO, I have argued elsewhere that the actual WTO does not seem well suited to the task. First, as mentioned, the WTO seems incapable of getting beyond the Doha Round, and incapable of addressing deep disciplines until it does. Second, a situation where China, India and other large emerging market stay outside the megaregionals may prove to be stable. The 'soft preferences' arising from the megaregionals may not prove very damaging to large outsiders who can use their market size and unilateral harmonization to offset the negative effects. This is what happened in Europe in reaction to the Single Market integration. Most outsiders joined the EU, but a few, like Switzerland and Norway, chose to stay out and make adjustments to live with the soft preferences. A domino effect, however, is very likely to draw in smaller outsiders wishing to participate in the international production networks inside the megaregionals.

What all this suggests is that world trade governance is heading towards a two pillar system. The first is the WTO which continues to govern traditional trade as it has done since it was founded in 1995. The second is a system where disciplines on trade in intermediate goods and services, investment and intellectual property protection, capital flows, and the movement of key personnel are multilateralised in megaregionals. China and certain other large emerging markets may have enough economic clout to counter the exclusion from megaregionals. Live and let live within this two-pillar system is a very likely outcome.

References

- Acharya, Rohini, Jo-Ann Crawford, Maryla Maliszewska and Christelle Renard (2011). "Landscape", *Preferential Trade Agreement Policies for Development: A Handbook*, Jean-Pierre Chauffour and Jean-Christophe Maur (eds), World Bank, Washington DC.
- Bagwell, Kyle and Robert Staiger (2004). *The Economics of the World Trading System*, MIT Press.
- Baldwin, Richard (2010). "Understanding the GATT's wins and the WTO's woes", CEPR Policy Insight No.49.
- Baldwin, Richard (2011). "Trade and Industrialisation after Globalisation's 2nd Unbundling: How Building and Joining a Supply Chain are Different and Why it Matters," NBER WP 17716.
- Baldwin, Richard (2012). "WTO 2.0: Global governance of supply-chain trade." CPER Policy Insight 64.
- Baldwin, Richard (2014). 'Multilateralising 21st Century Regionalism,' OECD Global Forum on Trade Reconciling Regionalism And Multilateralism In A Post-Bali World, February 2014.
- Baldwin, Robert (1970). *Nontariff Distortions of International Trade*. The Brookings Institution, Washington.
- Baldwin, Robert (2009). "Trade negotiations within the GATT/WTO framework: A survey of successes and failures," *Journal of Policy Modeling*, vol. 31(4), pages 515-525.

- Bhagwati, Jagdish (1993). "The Diminished Giant Syndrome", *Foreign Affairs*, pp 22-26.
- Clemens, Michael and Jeffrey Williamson (2004). "Why Did the Tariff-Growth Correlation Change after 1950?" *Journal of Economic Growth* 9 (1): 5-46.
- Croome, John (1996). *Reshaping the world trading system: a history of the Uruguay Round*. DIANE Publishing.
- Estevadeordal, Antoni, Caroline Freund and Emanuel Ornelas (2008). "Does Regionalism Affect Trade Liberalization toward NonMembers?," *The Quarterly Journal of Economics*, vol. 123(4), pages 1531-1575.
- Ethier, Wilfred (1998). "Regionalism in a Multilateral World," *Journal of Political Economy*, vol. 106(6), pages 1214-1245.
- Grossman, Gene and Elhanan Helpman (1995). "Trade Wars and Trade Talks", *Journal of Political Economy*, Vol. 103, No. 4, pp. 675-708.
- Horn, Henrik, Petros C. Mavroidis, and André Sapir (2010). "Beyond the WTO? An anatomy of EU and US preferential trade agreements." *The World Economy* 33, no. 11: 1565-1588.
- Irwin, Douglas A., Petros C. Mavroidis, and Alan O. Sykes (2008). *The Genesis of the GATT (The American Law Institute Reporters Studies on WTO Law)*. Cambridge University Press.
- Hoekman, Bernhard and Michael Kostecki (2010). *The Political Economy of the World Trading System: From GATT to WTO, Third Edition*, Oxford University Press.
- Kenen, Peter B (1989). *Exchange rates and policy coordination*. Manchester University Press.
- Keohane, Robert and Joseph Nye (2001). *Power and Interdependence*. 3rd edition: New York: Addison-Wesley Longman.
- Lawrence, Robert Z (1996). *Single world, divided nations?: International trade and OECD labor markets*. Brookings Institution Press.
- Lockwood Ben and Ben Zissimos (2004). "The GATT and gradualism", mimeo.
- Oyejide, Ademola (2002). "Special and differential treatment", in B. Hoekman, A. Mattoo, and P. English, *Development, Trade and the WTO*, World Bank.
- Patel, Mayur (2007). "New faces in the green room: developing country coalitions and decision-making in the WTO", *Global Economic Governance Programme Working Paper* 33.
- Preeg, Ernest (1970). *Traders and diplomats: An analysis of the Kennedy Round of Negotiations under the GATT*, Brookings Institution.
- Putnam, Robert (1988). "Diplomacy and Domestic Politics: The Logic of Two-Level Games". *International Organization* 42: 427-460.
- WTO (2011). "The WTO and preferential trade agreements: From co-existence to coherence", *World Trade Report*, 2011, WTO Secretariat.